Developing complexity using networks of synthetic replicators

Tamara Kosikova

University of
St Andrews

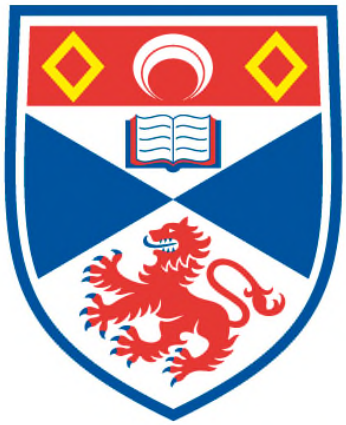
This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy
at the University of St Andrews

November 2016

DEVELOPING COMPLEXITY USING NETWORKS OF SYNTHETIC REPLICATORS

Tamara Kosikova

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2016

1. Candidate's declaration:

I, Tamara Kosikova, hereby certify that this thesis, which is approximately 80000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2012 and as a candidate for the degree of Doctor of Philosophy in September 2013; the higher study for which this is a record was carried out in the University of St Andrews between 2012 and 2016.

Date ..... signature of candidate .....

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ..... signature of supervisor .....

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all printed copies but an embargo on all electronic copies for a period of 1 year on the following ground: publication of this thesis would preclude future publications.

Date ..... signature of candidate ..... signature of supervisor .....

CONTENTS

ACKNOWLEDGEMENTS

The last four years have been a tremendous adventure—an adventure that was made significantly more enjoyable by the contributions of the people that helped me through it. First and foremost, I would like to thank my supervisor Professor Douglas Philp—for letting me loose in his lab, for his invaluable help with the challenges that occasionally accompany research in the field of systems chemistry, his incredible enthusiasm, and for what he might call "being picky", but really, it means doing things the right way.

I had the privilege of sharing my PhD journey with an amazing group and lab mates. In the Philp group, Josh Richards and Leo Maugeri (congratulations Dr!) started their PhDs almost at the same time as I did, and together with David van Brussel, we had the unique chance to build on the great work produced by the previous members of the Philp lab, explore the wonders of self-replicating systems and get really good at synthesising all things insoluble. Thank you guys for all the discussions, the jokes, the company and the outings. During my PhD, I also had the opportunity to work with a number of great project students—thank you guys for the hard work and for making the lab such a lively, gently crazy place. I would like to thank Martin Peeks in particular, who worked on his project during my first year: the discussions were invaluable and I shall forever remain impressed by your skills with all things computer-related. Further, I want to thank the members of the Kay research group: Dr Euan Kay for his extremely helpful discussions and comments; Dr Flavio della Sala for his Italian ways and for his help and patience when I first joined the lab; Dr Stefan Borsley for baking us all those cakes (while not eating any!) and most importantly for his friendship both inside and outside of the lab. Nicolas Marro—thank you for being such a cheerful presence in the lab!

I am grateful to the DoH group, particularly Rudy, Tanya, Dr Ricardo and Rodrigo for livening up the 4th floor corridor and for joining us on our Friday pub adventures. Special thanks go to Dr Maciek (Squirrel) Skibinski, for being an excellent friend, a

fellow fan of Carcassonne and cake, and for taking me on my very first trip to the Highlands (which was followed by many more). I would also like to thank Dr Stephen Thompson, for being an overall amazing and kind person and for introducing me to the loveliest Silvie in the world. Dr Neil Keddie, thank you for all the coffee chats, they were an often-needed break from the everyday craziness, and the expert discussions and assistance with all things chemistry.

The work done to produce this body of work has benefited tremendously from the help of the technical staff who work here at the School of Chemistry. I would like to thank Mrs Melanja Smith for brightening my visits to the NMR room and for her shared love of ceramics. Thank you and Dr Tomáš Lébl for your kind help with the numerous NMR experiments and for teaching me how to run them on my own. Mrs Caroline Horsburgh—thank you for the numerous mass spectrometry analyses; great thanks also go to Dr Matt Fuszard and Dr Sally Shirran for the MALDI-TOF analysis of my rotaxanes. I feel tremendous gratitude for the wonderful X-ray analysis performed by Dr David B. Cordes and Professor Alexandra M. Z. Slawin.

To my friends back home—Lucka, Dee, Mata, Matus and Amalka—thank you for all your love, our shared trips and the occasionally late-night chats. The number of wonderful people that I have met here in St Andrews is greater than I can mention—thank you all for making St Andrews the great place to be that it is. A special thank you goes to Lisa—you are an inspiration and who knows, maybe one day we will open that Café. To JoJo, the man I love*—thank you for sharing this journey with me.

LaTeX—thank you for making the whole thesis writing processes significantly more interesting and, hopefully, just a little bit easier than it would have been otherwise.

In the end, I would like to thank my family and my parents in particular—mami and tati, you have both been so incredibly wonderful and supportive during my time here—thank you for all your love and unfaltering support.

ABBREVIATIONS

aq.	aqueous
br	broad
BZ	Belousov-Zhabotinsky
C	cytosine
CA	Carbonic anhydrase
CI	chemical ionisation
CSTR	continuously-stirred tank reactor
d	doublet
dd	doublet of doublets
decomp.	decomposition
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCC	dynamic covalent chemistry
DCL	dynamic covalent library
DFT	density functional theory
DKS	dynamic kinetic stability
DNA	deoxyribonucleic acid
DOSY	diffusion-ordered spectroscopy
ε	autocatalytic efficiency
ee	enantiomeric excess
$\mathbf{EM_{kinetic}}$	kinetic effective molarity
$\mathbf{EM_{thermo}}$	thermodynamic effective molarity
ESI	electrospray ionisation
FMO	frontier molecular orbitals
$\mathbf{\Delta G^S}$	free energy of connection
G	guanine
GM	glycol macrocycle
h	hour(s)

HOMO	highest occupied molecular orbital
HRMS	high resolution mass spectrometry
ILG	informational leaving group
$k_{\mathbf{bi}}$	bimolecular reaction rate constant
$k_{\mathbf{uni}}$	unimolecular reaction rate constant
$K_{\mathbf{a}}$	association constant
$K_{\mathbf{d}}$	dissociation constant
LG	leaving group
lit.	literature
LUCA	Last Universal Common Ancestor
LUMO	lowest unoccupied molecular orbital
m	multiplet
M	maleimide
M.p.	melting point
MALDI	matrix-assisted laser desorption ionisation
Me	methyl
MIA	mechanically-interlocked architecture
MS	mass spectrometry
NBS	*N*-bromosuccinimide
NMR	Nuclear magnetic resonance
p	dissociation dependent autocatalytic reaction order
PhMe	toluene
PM	pyridine macrocycle
ppm	parts per million
***p*TSA**	*para*-toluenesulfonic acid
quant.	quantitative
R	rotaxane
RR	reciprocally-replicating
RM1	recife model 1
RMS	residual fitting error
RNA	ribonucleic acid
rOe	rotating frame Overhauser effect
ROESY	rotating frame nuclear Overhauser effect correlation spectroscopy
RR	reciprocal replicator
RT	room temperature
s	singlet
sat.	saturated
SR	self-replicator / self-replicating reactivity
t	triplet
t	time

T	template or thread
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TMS	trimethylsilane
TS	transition state
TSA	transition state analogue
UV	ultraviolet
WSBR	well-stirred batch reactor
wt	weight

ABSTRACT

Molecular recognition plays an essential role in the self-assembly and self-organisation of biological and chemical systems alike—allowing individual components to form complex interconnected networks. Within these systems, the nature of the recognition and reactive processes determines their functional and structural properties, and even small changes in their identity or orientation can exert a dramatic effect on the observed properties. The rapidly developing field of systems chemistry aims to move away from the established paradigm in which molecules are studied in isolation, towards the study of networks of molecules that interact and react with each other. Taking inspiration from complex natural systems, where recognition processes never operate in isolation, systems chemistry aims to study chemical networks with the view to examining the system-level properties that arise from the interactions and reactions between the components within these systems.

The work presented in this thesis aims to advance the nascent field of systems chemistry by bringing together small organic molecules that can react and interact together to form interconnected networks, exhibiting complex behaviour, such as self-replication, as a result. Three simple building blocks are used to construct a network of two structurally similar replicators and their kinetic behaviour is probed through a comprehensive kinetic analysis. The selectivity for one of the recognition-mediated reactive processes over another is examined within the network in isolation as well as in a scenario where the network is embedded within a pool of exchanging components. The interconnected, two-replicator network is examined under far-from-equilibrium reaction-diffusion conditions, showing that chemical replicating networks can exhibit signs of selective replication—a complex phenomenon normally associated with biological systems. Finally, a design of a well-characterised replicator is exploited for the construction of a network integrating self-replication with a another recognition-directed process, leading to the formation of a mechanically-interlocked architecture—a [2]rotaxane.

CHAPTER 1

INTRODUCTION

1.1 Preamble

The world around us has come to rely on networks: social networks,[1,2] citation networks,[3,4] transportation networks,[5] the world wide web,[6] and, in fact the human body itself is a vast network comprised of a large number of smaller interconnected neural, and gene signalling networks[7,8]—these are just few examples of the networks that we experience in our everyday lives. These networks differ greatly, ranging in the number of components, the links between them, and their strength, whether the networks are physically tangible (*e.g.* transportation network) or more abstract (*e.g.* friendship networks). One feature, however, is shared by all of these networks, and that is the dramatic influence they exert on our lives:

> "August 2003: A massive power outage hit the Midwestern and Northeastern United States, caused by a cascading failure due to a shutdown at one generating plant in Ohio. The reported cause of the shutdown was that electrical lines, overloaded by high demand on a very hot day, sagged too far down and came into contact with overgrown trees, triggering an automatic shutdown of the lines, whose load had to be shifted to other parts of the electrical network, which themselves became overloaded and shut down. This pattern of overloading and subsequent shutdown spread rapidly, eventually resulting in about 50 million customers in the Eastern United States and Canada losing electricity, some for more than three days."[9,10]

Traditionally, the study of networks was the domain of mathematics, referred to as Graph Theory, as pioneered[11,12] by Euler in the early 1740s. Over time, the study

of networks has achieved significant importance, particularly in the field of social sciences,[13] with a focus on unravelling communication and relationships between individuals as well as larger social units. The development of graph theory has allowed networks from the real-world to be represented in the form of graphs, where the graph vertices (or nodes) denote the components or individuals within a particular network, and the edges represent the connections between them (**Figure 1.1a**). A network can incorporate different types of nodes and edges **Figure 1.1b** illustrates an example of a network with directed edges (*i.e*. a network where the interactions between nodes are directional), such as in a network representing telephone calls. Alternatively, the edges can carry also different weights, which can be represented by the thickness of the edge line—a property of an edge that can vary greatly, for example, in friendship networks (**Figure 1.1c**).

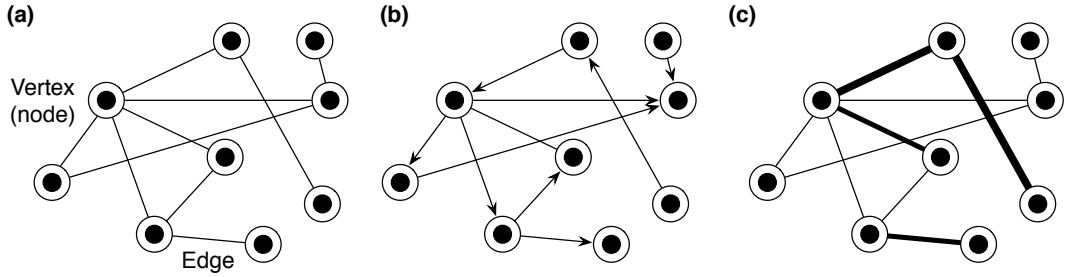


Figure 1.1 A depiction of a simple **(a)** undirected, **(b)** directed and **(c)** undirected weighted network comprised of nine identical vertices (nodes) and ten connections (edges). Figure adapted from Ref. 12.

The representation of a network in the form a graph relays information about the structure of the network and its topology, which in turn provides information about its function and behaviour. For example, a graph representing a social network can be analysed in terms of the connectivity exhibited by various nodes, in particular the number (the degree of connectivity) and the strength of edges from a particular node, revealing the most essential nodes for the network to retain its function. However, extracting structural and functional information from the graphical representation of a network becomes significantly more difficult, if not impossible, as the number of components in a network increases beyond a certain critical threshold. Therefore, visual analysis that can provide facile access to information about smaller networks (**Figure 1.2c**), becomes less important for large networks (**Figure 1.2a** and **Figure 1.2b**).

Following the appearance[7,13] of two seminal papers, discussing the collective dynamics of the so-called 'small-world' networks, the research directed at investigations of complex networks has experienced a renewed surge across a large variety of disciplines. In this context, the term complex network is used[12] to represent networks that increase in complexity not simply as a result of the growing number of a components,

Figure 1.2 Three example networks varying significantly in the number of components (nodes) and interconnections (edges). **(a)** The structure of the internet network, where each node represents a collection of computers. **(b)** A predator-prey network showing the interactions among the species inhabiting a single lake and **(c)** a social network of sexual contacts. Figure taken from Ref. 11.

but, instead, through the dynamic nature of the links themselves, which enable the networks to evolve over time. In complex networks, the degree of interconnectedness between the individual components can change over time, and an alteration in one network component can exert a dramatic influence on the rest of the network, resulting in what has become known popularly as the Butterfly effect[14] in chaos theory—a term coined initially to denote systems that are highly sensitive to their initial conditions. The concept of feedback is crucial in complex systems, with the potential to amplify or dampen the effect of a particular change or stimulus.

Interestingly, research directed at complex networks has revealed[13] that most of the identified dynamic networks in the real world share certain identifying features. Namely, these complex networks, irrespective of their nature (biological, social, etc.), are neither completely regular (**Figure 1.3a**) nor completely random (**Figure 1.3c**) in terms of edge distribution, but lie instead somewhere in between these two extremes (**Figure 1.3b**). Watts and Strogatz coined[13] the term 'small-world' networks[15,16] to describe such systems, using the analogy to the small-world phenomenon, perhaps better known as six-degrees of separation.[17] Complex networks also contain highly interconnected sections (known as clusters) and exhibit short average path length, where path length is understood as the shortest path between any two nodes. The example of small-world network illustrated in **Figure 1.3** demonstrates this tendency of 'small-world' networks to exhibit reduced path length, generally caused by few random links between nodes—connections that are absent in regular networks. The short path length and clustering properties of small-world networks have been hypothesised[9] to have

emerged as a result of evolutionary pressure for fast information transfer (small path-length) and challenges associated with maintaining links over long-distances (only few random links are necessary to decrease path length significantly).

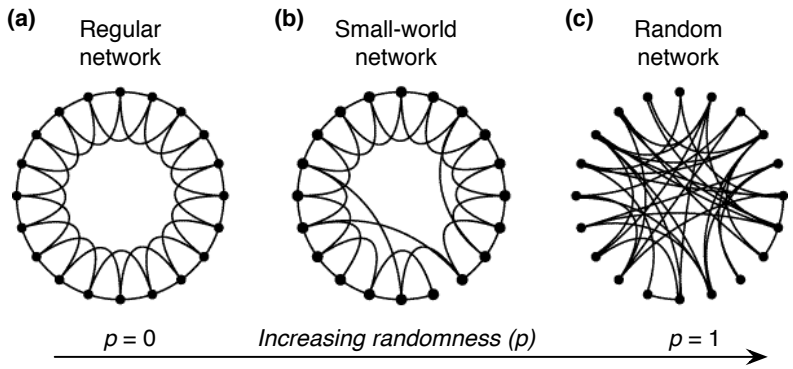


Figure 1.3 Three example networks comprised of 20 nodes, illustrating a **(a)** completely regular, **(b)** small-world and **(c)** random distribution of edges (interconnections). Probability *p* represents the likelihood of an edge, associated with a particular node, of being randomly wired. Figure taken with permission from Ref. 13. Copyright 1998 Nature Publishing Group.

The structure of a network, its components and their interactions determine the function and behaviour of a network. Analysis of how structural features and connections affect complex networks, and the use of this knowledge in simulations, can facilitate predictions about the development of a complex network over time to be made. The topology of a network has been found to play a crucial role in determining whether or not a network can expresses[18] emergent properties, such as synchronisation and robustness. An emergent phenomenon is one that is the result of a global expression of system-level interactions, *i.e*. it does not stem from the properties of the individual components in isolation, but rather arises from their interactions. Biochemical networks are a great example of complex networks where individual molecules come together to create responsive networks, capable of signalling, gene regulation, and many other functions—with the capacity to adapt and evolve over time in response to changes in the environment and input. Similarly, the wiring of social networks can have, for example, tremendous implications for the spread of infections diseases. As a result of the common underlying principles found throughout complex dynamic systems, research directed at examining complex networks has the potential to improve greatly our understanding of these systems as well as the world in general—from the origin of emergent phenomena to how infectious diseases spread through populations.

1.2 Origins of life

Life can likewise be considered as an emergent property that arises from the interactions of a large number of enzymes, proteins, oligonucleotides, membranes and many others components, that would not be considered alive in isolation otherwise. The examination of the living complexity of life brings[19,20] to mind the question—how did life originate? What was the sequence of events that lead to the transition from a prebiotic world filled with simple molecules to a world abounding with living organisms? The origin of life has been[21] the subject of scientific enquiry for centuries and the subject of man's enquiry for even longer.

Perhaps the earliest belief associated with the emergence of living organisms is that of spontaneous generation—a belief that life arises naturally from inanimate matter. After a prolonged debate, the matter of spontaneous emergence was finally settled by Pasteur around 1860s, using a series of simple and elegant experiments. Pasteur developed a swan-necked flask that possessed an S-shaped neck, which prevented any micro-organisms from reaching the sterile solution contained within. Once the neck was broken off, the bacteria gained access to the previously sterile broth, allowing the bacterial populations to multiply and, thus, demonstrating that life emerges only from other living matter. Whilst an essential step towards the understanding of living systems, the question of where and how life originated remained unanswered. At the beginning of the 19$^{\text{th}}$ century, the vitalist view, advocating that living organisms possess a special non-physical, but further unspecified quality that fundamentally distinguishes them from non-living entities—the vital force, was dominant. Only with the demonstration that an organic compound, namely urea, can be prepared in the laboratory "without the need for an animal kidney"[22] by Wöhler, did the vitalist view begin to slowly lose credibility. This first example of a synthesis of an organic product using purely inorganic starting material served as a bridge between the living and non-living world, whilst simultaneously heralding the beginning of a new era of chemical research and synthetic chemistry.

The understanding of living organisms was revolutionised by Darwin's theory on the Origin of Species,[23] with three main processes at its core driving evolution: replication, mutation and selectivity. Over time, further understanding into the specific nature of living organisms was provided with the advent of molecular biology in the 20$^{\text{th}}$ century. Replication, in the most general terms, involves the formation of an identical copy of an entity from a single original unit and is a complex phenomenon that is ubiquitous in Nature. The capacity for replication in biological systems emerges[24,25] not simply as a function of the large number of components present within a cell. Instead, it develops[18,26] because the constituent components of these systems have the capacity

to interact and react with each other, forming dynamic, self-organised, interconnected networks. It is these complex networks, in particular, that give rise[8] to the emergent behaviour of life.

In 1953, Watson and Crick have, assisted by the work of Franklin and Wilkins, uncovered[27,28] the double helical structure of deoxyribonucleic acid (DNA), revealing the template-directed mechanism that enables the copying and transmission of the genetic material of an organism. This work highlighted the reliance of the replication mechanism on the specific molecular recognition between nucleotide sequences of the informational biopolymer. During replication, two helical DNA strands act as individual templates for the synthesis of one new DNA strand, producing sequences with nucleotide pattern that is complementary to each parent template strand. While the mechanism for replication is conceptually simple, the assistance of numerous components of the cellular machinery, such as enzymes, is required in order for replication to take place. Ultimately, DNA, together with ribonucleic acid (RNA) and proteins lie[29] at the very core of biology, encoding within their structure the instructions that enable living organisms to function and replicate persistently. However, in spite of the richness of detail now available for nearly every cellular process, the answer to the question "what makes a cell alive?" remains unclear. As Albert Eschenmoser aptly noted[30] "life as we know it is a chemical life". This means that it should be possible to explain the biological processes making life possible, in terms of the underlying chemistry.

The structural features and uniformity of the current genetic material suggest[31] that the modern living systems emerged[32,33] from a single Last Universal Common Ancestor (LUCA), marking the transition from an inanimate world to a recognisably living entity. However, our understanding and agreement as to what may constitute[19,34–37] life remains controversial. NASA has provided[38] the most generally accepted definition of life, where "a self-sustaining chemical system capable of undergoing Darwinian evolution" is considered to be living. We are unlikely to ever identify the concrete steps that facilitated the progressive spontaneous increase in complexity on Earth. The transformation to a living world, filled with a biodiversity of living organisms, required that the prebiotic soup of simple, chemical components developed into larger functional biopolymers. These functional materials must have been capable of catalytic activity and self-replication, and have possessed the capacity to couple with primitive metabolic cycles and membrane-based compartments. In the 19th century, Darwin's theory of evolution transformed our understanding of how populations of species adapt and evolve over time through repeated cycles of replication, mutation and selection. The ability of molecules to self-replicate serves as a bridging element between chemistry and biology. Recently, it has been suggested[39–41] that the principles at the core of Darwinian

theory can be extended to the physiochemical processes leading to LUCA. The prebiotic chemistry literature abounds[21,30,42,43] with numerous models and theories that strive to explain how such gradual transition could have taken place. These models generally fall into one of three main categories, differing in their approach to the experimental study of life's origins: RNA world hypothesis,[44–47] metabolism first hypothesis[48–50] and the compartmentalistic approach.[51–53] These schools of thought tend to disagree on the identity of the molecular species believed to have emerged first during the process of chemical evolution. All three frameworks, however, are united in their belief that autocatalysis has played a crucial role as the predecessor of modern self-replication in the sequence of events leading to the emergence of the modern biodiversity. Recently, integrative systems approaches, where individual monomers and larger oligomers and assemblies, primitive metabolic networks, catalytic species, metals and minerals all co-evolved simultaneously within the environment of the aqueous prebiotic soup are beginning to replace[54,55] the more traditional, less holistic theories.

1.3 Systems chemistry

Undeniably, life and its origins are linked with the chemistry of complex systems. In general, however, the recognition of the role of complexity in both biology and chemistry has been slow to develop,[56,57] and lagging behind other science disciplines. Recently, though, new systems-based approaches have emerged, first in the field of biology,[58–60] and later also in chemistry.[43,61–63] These approaches are directed at trying to understand chemical and biological complexity using a 'bottom-up' approach. In particular, systems chemistry aims[64–66] to explore the connections between the properties of individual components (molecules) and the emergence of complex, dynamic, system-level behaviour as a result of the interactions of these components. In exploiting synthetic chemistry for the design and development of systems with complex and potentially life-like properties, systems chemistry strives to develop a better understanding of the principles governing assembly and function in complex systems, thereby shedding light on the origins of biological complexity.

A hypothetical and structurally-simple example of a chemical reaction network can be imagined in the form of reactants interconnected by a shared building block. **Figure 1.4a** shows the topology of a three component network that can result in the formation of two products: **AB** and **AC**, linked by a common building block **B**. In the scenario where the two reactions have the same rate constant ($k_{AB} = k_{BC}$) and all three components are present at an identical concentration, **AB** and **BC** will reach the same concentration once **B** is fully consumed (**Figure 1.4a**). However, if the rate constant for

one of these reactions is upregulated (increased), the product of this pathway will reach a higher concentration. The example for $k_{AB} > k_{AC}$ is illustrated in **Figure 1.4b**.

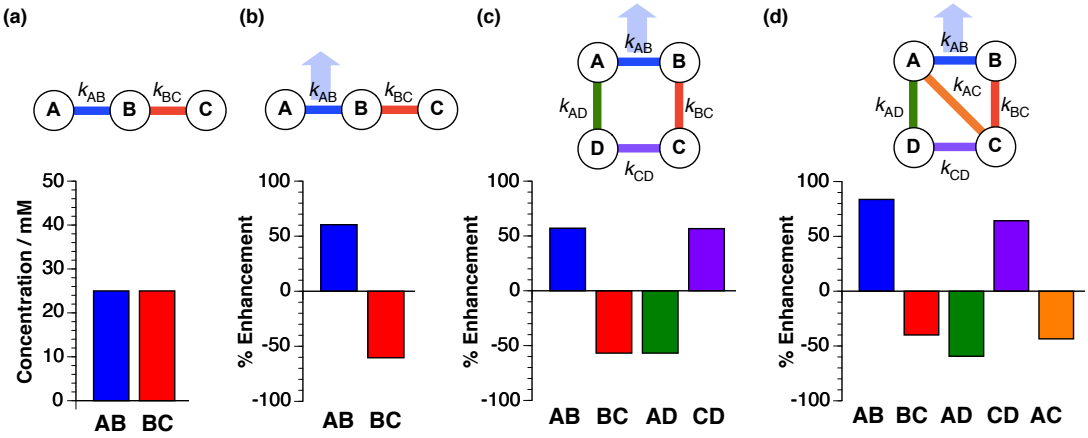


Figure 1.4 Simulation outcome for a series of simple reaction networks composed of three to four components. Circles represent reaction building blocks, whilst coloured arrows denote the formation of product, governed by a specific reaction rate constant. Initial concentration of all components was 50 mM and each rate constant was set at 10^{-3} $M^{-1}s^{-1}$. Blue arrows represent a 10-fold increase in the rate constant k_{AB} to 10^{-2} $M^{-1}s^{-1}$ in conditions **(b)** through **(d)**. Product distribution was determined after *ca.* 30 h.

Building on these two straightforward examples of simple reaction networks, the system can be extended to incorporate an additional building block, **D**, affording a network with four different products: **AB**, **BC**, **CD** and **AD**, each formed with a specific rate constant (**Figure 1.4c**)—resulting in a network with a square topology. If the rate constants are identical, the components will be equally distributed across the four products (data not shown). An increase in one of the rate constants, k_{AB} for example, relative to the remaining rate constants, will result in an increased concentration of **AB** product. Interestingly, however, the increase in the rate constant for the formation of **AB** will also result in the concentration of **CD** being higher than the concentration of **BC** and **AD**. The **CD** product, whilst formed at a slower rate than **AB**, will reach the same concentration as **AB** over time. The fast consumption of the reactants **A** and **B** means that the amount of these components available for reaction with components **C** and **D** decreases relative to the condition where all four rate constants are identical. As a result of this imbalance, the two components that are not consumed efficiently (**C** and **D**) remain available in the reaction mixture, where they can react together slowly over time. Hence, it is possible to observe behaviour that can be considered emergent with respect to the upregulation of component **CD**, afforded solely through the interactions in the system. If this four-component network is extended by another edge, between **A** and **C** (**Figure 1.4d**), the effect of increasing one rate constant becomes yet more interesting. It can be envisaged that these simple examples of chemical reaction networks can increase significantly in complexity, if the reaction network incorporated molecules

capable of recognition-mediated interactions and reactions and feedback, for example, networks involving molecules that are capable of catalysing their own formation, *i.e.* self-replication.

Cellular systems normally possess much more intricate networks, where individual interactions are not easily decoupled. By examining the dynamic processes governing replication in synthetic systems, systems chemists are able to acquire a better understanding of complex phenomena since the networks involved can introduce interactions between components in a designed and organised manner. Moreover, through the study of synthetic reaction networks capable of replication, it is possible to arrive at a set of minimal structural, recognition and reaction requirements necessary for a system to be capable of replication, delineating boundary conditions for the emergence of the enzymatic machinery available to modern cells.

Ultimately, the potential of systems chemistry to bridge the gap between the chemical and the biological world is facilitated by advances in the tools available for the analysis of multi-component systems used to study complex phenomena, by combing results from experimental systems with computational simulations. In the last two decades, significant progress has been made[67–70] in the design and experimental demonstration of minimal *in vitro* self-replicating systems, based on biological molecules and organic molecule alternatives, both in isolation and within more complex interconnected networks.

1.4 Replication: theory and practice

1.4.1 Minimal model of self-replication

A system capable of self-replication is one that can transmit structural information through an autocatalytic process. What distinguishes self-replication from the larger subset of autocatalytic reactions is the specificity with which the product formed in the autocatalytic reaction acts as a specific catalyst for its own formation from the individual building blocks. For example, acid-catalysed ester hydrolysis results in the formation of a product that is in itself an acid—accelerating further the hydrolysis reaction. This reaction, however, is non-specific, and will be accelerated by any acid, not just the product itself. In a self-replicating system, the product itself acts as a specific catalyst for its own formation. Consequently, the rate of self-replicating reactions correlates directly with the amount of product template present within a particular reaction mixture.

Within the minimal model of self-replication[71] (**Figure 1.5**), product formation can proceed through three different reaction channels. In this model, two molecules, **A** and

B, are equipped with complementary reactive (orange and green) and recognition sites (yellow and blue). In Channel 1 (**Figure 1.5**, black), these two building blocks react through a slow and uncatalysed template-independent pathway, affording template $\mathbf{T^{AB}}$ or $\mathbf{T^{AB\star}}$. Alternatively, the same molecules can react through two recognition-mediated channels. The complementary recognition sites on **A** and **B** enable association of these components in a binary complex [**A**·**B**]. Formation of the binary complex in Channel 2 (**Figure 1.5**, red) facilitates the reaction between **A** and **B** by preorganising their reactive sites, resulting in the formation of a catalytically-inactive, closed template $\mathbf{T^{AB\star}}$.

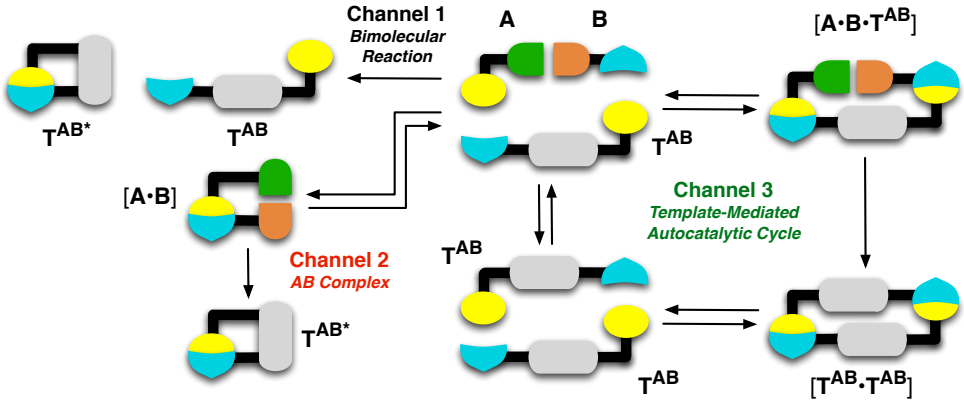


Figure 1.5 Cartoon representation of the minimal model of self-replication. Components **A** and **B** are equipped with a reactive (green/orange) and recognition site (yellow/blue). Formation of template $\mathbf{T^{AB}}$ can proceed through slow bimolecular reaction (Channel 1, black) and template-mediated self-replicating pathway (Channel 3, green). Channel 2 (red) describes the formation of a closed template $\mathbf{T^{AB\star}}$ through a binary reactive complex [**A**·**B**] complex.

Ultimately, the **A** and **B** components can react *via* a template-catalysed recognition-mediated pathway, referred to as Channel 3 (**Figure 1.5**, green). In this pathway, the template $\mathbf{T^{AB}}$, formed through the bimolecular reaction initially, possesses recognition-sites in an 'open' conformation that allow it to assemble with the unreacted components **A** and **B** in a catalytically-active ternary complex [**A**·**B**·$\mathbf{T^{AB}}$]. Within this complex the reactive sites present on **A** and **B** are preorganised for reaction, generating rate enhancement in what is effectively a pseudounimolecular reaction. The outcome of the template-mediated reaction is a template dimer [$\mathbf{T^{AB}}$·$\mathbf{T^{AB}}$], which supplies the reaction mixture with two identical molecules of template, with the capacity to take part in further autocatalytic cycles.

Self-replicating systems often present with a sigmoidal reaction profile. The shape of the reaction profile results from the fact that the reaction rate within a self-replicating system in a closed reaction environment depends on two main factors: (i) the concentration of the free catalytically-active (open) template and (ii) the availability (*i.e.* concentration) of the starting materials. The lag phase at the beginning of a reaction pro-

file is present typically because the building blocks have to first react through the slow bimolecular reaction to produce the catalytically-active template before the reaction can operate efficiently. The concentration of the template within a reaction mixture will continue to increase slowly, until such a time when the concentration is sufficient for association of the template with the reactants in the ternary complex. The quantity of template required for the assembly into catalytically-active components is governed by the dissociation constant K_d for the recognition-mediated processes in the corresponding system ($K_d = 1/K_a$). Nevertheless, once the template concentration passes the threshold, the predominant reaction pathway is the template-catalysed autocatalytic Channel 3—a phase in the reaction when the maximum reaction rate is observed in self-replicating systems. As a result of the direct dependence of the reaction rate on the concentration of free template, a decrease in the length of the lag period should be observed, proportional to the quantity of template added, if a self-replicating reaction is undertaken in the presence of preformed template (added at $t = 0$). In fact, kinetic experiments examining formation of replicators in the presence of preformed instructional template are utilised to unambiguously establish the capacity of a system to template its own synthesis. The autocatalytic pathway will continue to operate efficiently until the reaction components are exhausted—a time point at which the reaction profile transitions into the plateau phase.

1.4.2 Minimal model of reciprocal replication

In addition to the well-established and also the most studied minimal model of self-replication described above, replication processes can also operate in a reciprocal sense, *i.e.* by means of complementary recognition processes, instead of the self-complementary template required for self-replication. The somewhat less-studied minimal model of reciprocal replication (**Figure 1.6**) is comprised of four building blocks: **C**, **D**, **E** and **F**. From these four components, **C** can react with **D** to form template $\mathbf{T^{CD}}$, whilst the reaction of **E** and **F** produces $\mathbf{T^{EF}}$. These templates, formed through two independent, uncatalysed bimolecular pathways (Channel 1 and 2) posses recognition sites that are mutually complementary, and, therefore, do not permit self-association. As a result of the complementary nature of the recognition processes, template $\mathbf{T^{CD}}$ can associate **E** and **F** in a ternary catalytically-active complex [$\mathbf{E{\cdot}F{\cdot}T^{CD}}$] (Channel 3), thus accelerating the formation of template $\mathbf{T^{EF}}$. Similarly, $\mathbf{T^{EF}}$ catalyses the formation of $\mathbf{T^{CD}}$ *via* the formation of complex [$\mathbf{C{\cdot}D{\cdot}T^{EF}}$] (Channel 4). Dissociation of the resulting heterodimer [$\mathbf{T^{CD}{\cdot}T^{EF}}$] releases one molecule of each reciprocal template back into the reaction, enabling further catalytic cycles to proceed.

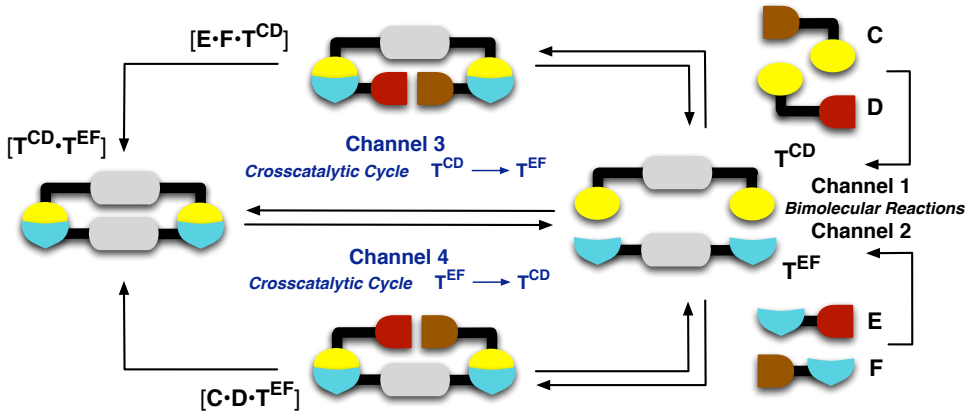


Figure 1.6 Cartoon representation of the minimal model of reciprocal-replication. Two templates, $\mathbf{T^{CD}}$ and $\mathbf{T^{EF}}$, containing complementary recognition sites (yellow and dark blue) are formed first formed by the bimolecular reaction of component **C** with **D** (Channel 1, black) and **E** with **F** (Channel 2, black). Once formed, the complementary nature of the recognition sites allows them to take part in template-mediated reciprocal replicating pathways, where $\mathbf{T^{CD}}$ is formed *via* the catalytically-active ternary complex [$\mathbf{C{\cdot}D{\cdot}T^{EF}}$] (Channel 4, blue) and $\mathbf{T^{EF}}$ *via* the analogous complex [$\mathbf{E{\cdot}F{\cdot}T^{CD}}$] (Channel 3, blue).

The formation of templates in a replication network operating in a reciprocal sense will be initially limited by the rates of the uncatalysed bimolecular reactions. Once the concentration of each reciprocal template is sufficient for the assembly of the reaction components in the catalytically-active ternary complexes, the template-directed pathways will begin to operate efficiently. The efficiency of the reciprocal system depends on the presence of free template molecules—replication trough the template-directed pathways can proceed only if the templates are not sequestered within template heterodimers.

1.4.3 Replication efficiency

The crucial role played by templates in both self- and reciprocal replication often makes these systems susceptible to product inhibition. In an ideal situation, the product duplex would be less stable than the catalytically-active ternary complex, thereby driving the dissociation of the dimer into individual molecules of free template. Synthetically, however, designing systems where the template dimer is less stable than the ternary complex is extremely challenging, and instead the template generally remains sequestered partially within product complexes to a smaller or larger degree, unavailable as the monomeric species required for efficient replication.

Very recently, a new theoretical model of self-replication aimed at reducing template inhibition has been reported[72] by Herdewijn and co-workers. Taking inspiration from the capacity[73] of non-coding RNA to perform function and their previous work on

nucleotide-based systems,[74] the authors propose a theoretical model of self-replication (**Figure 1.7a**) that exploits an informational leaving group (ILG) in order to produce weakly bound template duplexes and overcome product inhibition in replicating systems. In this ILG minimal model of self-replication, a template molecule $\mathbf{T^{GH}}$ is formed by the reaction of two precursors, **G** and **H** (Channel 1). Each precursor contains two recognition sites and one reactive site. Interestingly, the increased number of recognition sites means that the components **G** and **H** can theoretically associate together to form two binary complexes, [**G**·**H**]-1 and [**G**·**H**]-2 (**Figure 1.7b**) as well as longer chains.

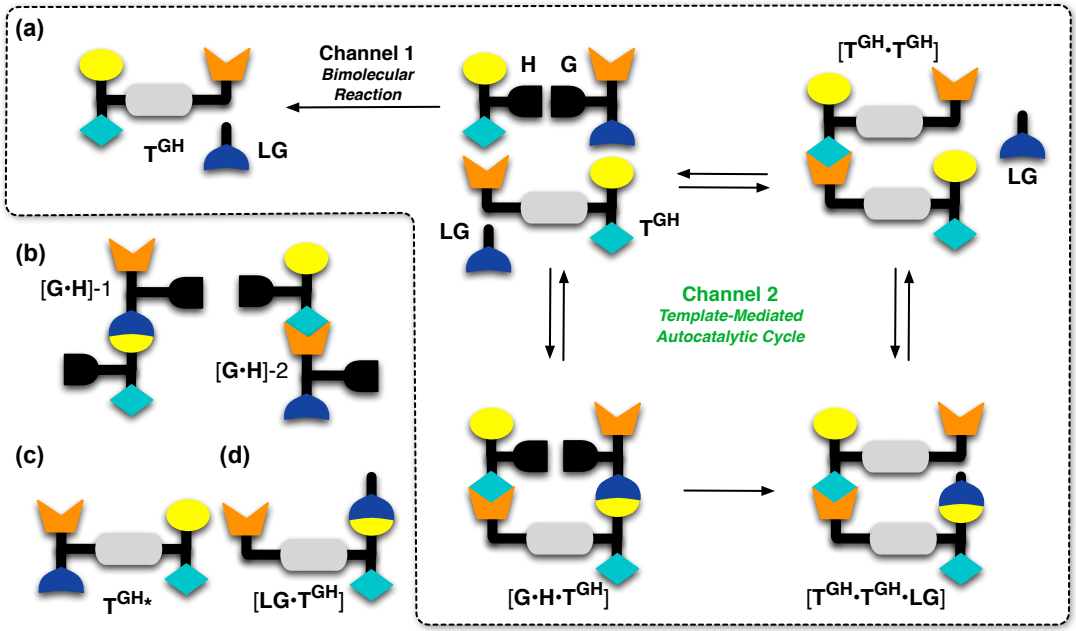


Figure 1.7 **(a)** Cartoon representation of a minimal model of self-replication exploiting an informational leaving group (ILG) strategy. Precursors **G** and **H** are each equipped with two recognition sites and a reactive site (black), react to form a molecule of template $\mathbf{T^{GH}}$. The reaction results in a release of the leaving group **LG** (Channel 1). The template $\mathbf{T^{GH}}$ can catalyse the reaction between the **G** and **H** by preorganising their reactive sites in a catalytically-active ternary complex $[\mathbf{G{\cdot}H{\cdot}T^{GH}}]$. The ligation step (Channel 2) produces a ternary complex containing two template molecules and a **LG**, $[\mathbf{T^{GH}{\cdot}T^{GH}{\cdot}LG}]$. Dissociation of the **LG** and the template returns two free catalytically-active molecules of $\mathbf{T^{GH}}$ into solution. **(b)** The recognition sites on the precursors **G** and **H** allow them to form two binary complexes [**G**·**H**]-1 and [**G**·**H**]-2. **(c)** Formation of template $\mathbf{T^{GH\star}}$ in the ILG model is not possible. **(d)** Association of **LG** with template $\mathbf{T^{GH}}$ in a complex $[\mathbf{T^{GH}{\cdot}LG}]$ can decrease catalytic efficiency. Figure adapted from Ref. 72.

The template forming reaction in this model is associated with a concomitant dissociation of the leaving group, **LG**, producing a template $\mathbf{T^{GH}}$ that is equipped with three recognition sites, as opposed to template $\mathbf{T^{GH\star}}$, the formation of which would be expected in the absence of the ILG (**Figure 1.7c**). Similarly to the minimal model of self-replication, the produced template $\mathbf{T^{GH}}$ can associate with the precursors, **G** and **H**, in a catalytically-active ternary complex $[\mathbf{G{\cdot}H{\cdot}T^{GH}}]$. The template preorganises the

reactive sites of these components, accelerating the formation of the ternary complex [$\mathbf{T^{GH}}$·$\mathbf{T^{GH}}$·**LG**] (**Figure 1.7a**, Channel 2). Because the formation of each template molecule is associated with the release of the leaving group **LG**, the resulting complex involves a molecule of template taking part in two individual recognition-mediated interactions with (i) another molecule of template and (ii) the leaving group **LG**, making the complex significantly less stable than the analogous template-dimer. Dissociation of **LG** from the ternary complex [$\mathbf{T^{GH}}$·$\mathbf{T^{GH}}$·**LG**] produces a template dimer [$\mathbf{T^{GH}}$·$\mathbf{T^{GH}}$], which is in equilibrium with the free-form of the catalytically-active template $\mathbf{T^{GH}}$. This novel theoretical ILG-based model of self-replication provides an interesting method for destabilising the product duplex, with the potential to improve replication efficiency, nevertheless, the problem of product inhibition, arising from the association of the free template $\mathbf{T^{GH}}$ with the leaving group **LG** in a complex [$\mathbf{T^{GH}}$·**LG**] is not resolved completely (**Figure 1.7d**), and the viability of this theoretical model awaits experimental testing.

In addition to exploring the use of an IGL, there are two reaction parameters can be varied fairly readily in order to improve the catalytic efficiency of a replicating system operating through one of the minimal models: concentration and temperature. By altering the concentration at which the reaction is performed, the proportion of free and bound template in solution can be manipulated. For example, changing the reaction concentration can be useful for controlling the relative contribution of the bimolecular and the recognition-mediated pathways towards the production of template. If, for example, the reaction concentration is halved, the bimolecular reaction will decrease by a factor of four, whereas the pseudounimolecular reaction will decrease only two-fold. Replicating systems generally contain a large number of various components present simultaneously in both bound and unbound state within the reaction mixture, and in order to deconvolute the overall effect of concentration on the reaction time course, it is often useful to employ kinetic simulations, utilising known kinetic and recognition parameters (*e.g.* rate and association constants).

Alterations in the experimental temperature can also be exploited for increasing the efficiency of replication. For example, an increase in the reaction temperature would result in a weaker duplex association, thereby decreasing product inhibition. Simultaneously, however, the temperature increase would affect also the formation of catalytically-active complexes. Changes in temperature influence at the same time both the reaction rate constants and the strength of recognition-mediated processes, and, therefore, the overall effect of changing temperature on the efficiency of replication can be challenging to predict. A decrease in the reaction temperature simultaneously slows down both the bimolecular and unimolecular reaction rate, whilst also increasing the

stabilities of all complexes relying on non-covalent interactions. The latter can perhaps lower the quantity of template required for its assembly with the reaction components in to catalytically-active complexes. Ultimately, though, each replicating system performs with the highest efficiency at a particular set of conditions.

1.4.4 Kinetic theory of self-replicating systems

Self-replication is a complex phenomenon, and as is often the case with complex behaviour, understanding it often necessitates simplifications. In 1993, von Kiedrowski reported[71] seminal work describing empirically-determined and simplified kinetic models for the kinetic behaviour of replicating systems. The simplest kinetic model (**Figure 1.8**) represents a purely autocatalytic system, *i.e.* a system where formation of template is possible only through a template-mediated pathway. Such self-replication process is characterised by an autocatalytic reaction order: p. Experimental replicating systems often fall between two extremes of reaction order, $p = 0.5$ and $p = 1$, values which correspond to parabolic and exponential growth, respectively. In a replicating system with $p = 0.5$, the initial rate of reaction follows the square-root law of autocatalysis, which means that the addition of increasing quantities of template results in an increase in the rate of template formation that is equal to the square root of the concentration of template added.

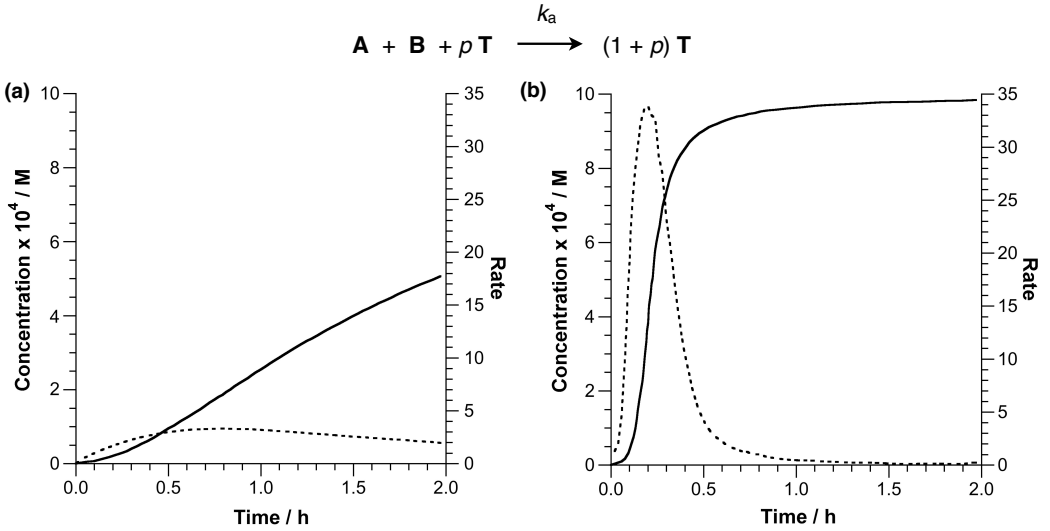


Figure 1.8 Concentration *vs* time (full line) and rate *vs* time (dashed line) profiles for a purely autocatalytic reaction with p equal to **(a)** 0.5 and **(b)** 1. Data taken from Ref. 71.

The magnitude of the reaction order p also conveys information about the rate-limiting step in the autocatalytic cycle. For $p = 0.5$, the dissociation of product duplex is the rate limiting step and such systems experience severe product inhibition. For

autocatalytic systems where the dissociation of product duplex is not the rate-limiting step, the reaction order p can be closer to 1. In exponential growth systems, where p = 1, doubling the amount of template added also doubles the rate of template synthesis.

The parabolic and exponential replicator growth curves portrayed in **Figure 1.8a** and **b** illustrate a purely autocatalytic system, and, therefore, these reaction profiles are not applicable to majority of experimental replicating systems. Instead, **Figure 1.9** shows the noticeably more varied behaviour obtained using a kinetic model that describes real systems more accurately. In this case, the product **T** is formed from **A** and **B** both *via* a template-independent (bimolecular) pathway and a template-directed (autocatalytic) pathway. Each pathway proceeds with a specific rate constant, k_a (autocatalytic) and k_b (bimolecular). The ratio of these two rate constants is referred to as the autocatalytic efficiency ε ($\varepsilon = k_a/k_b$). This parameter provides a measure of the contribution of each pathway to the overall concentration of template produced in a replicating system. Examining the parabolic and exponential time course profiles, it is clear that the kinetic behaviour varies significantly as a function of ε, and, in fact, the classical sigmoidal reaction profile is expressed only in replicating systems with ε above a certain threshold value. Therefore, in experimental systems that are thought to be self-replicating, but do not express a sigmoidal reaction profile as a result of low catalytic efficiency ε, the ability of a system to template its own synthesis can be confirmed indirectly, through template-instructed experiments.

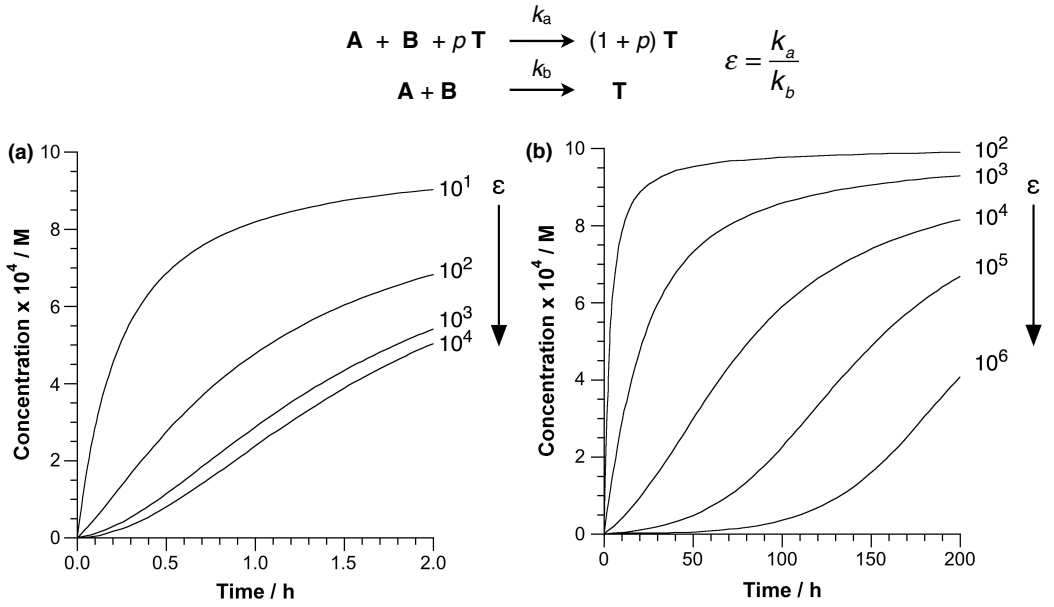


Figure 1.9 Effect of the ratio of k_a/k_b on the reaction profiles for more realistic autocatalytic reactions with p equal to **(a)** 0.5 and **(b)** 1. Data taken from Ref. 71.

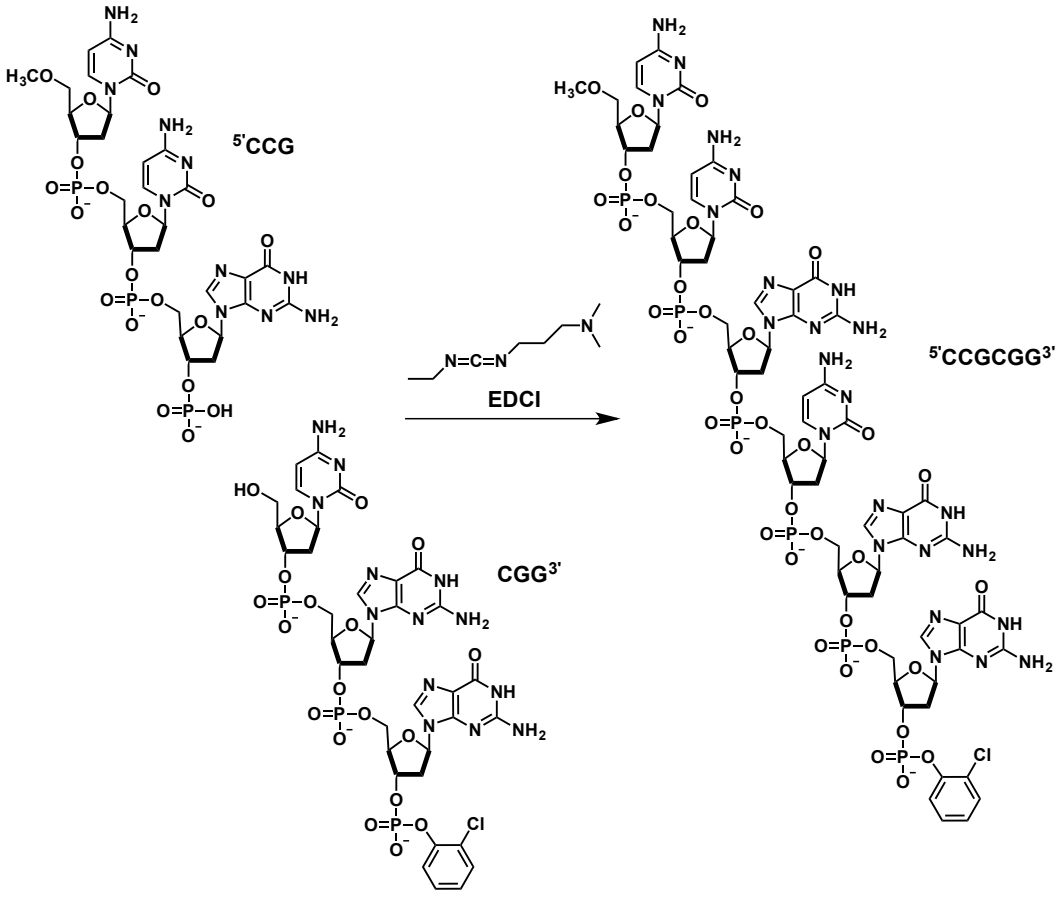
The large variation in reaction profiles that can be exhibited by experimental replicating systems demonstrates that ascertaining the rate-limiting step within each system as well as the catalytic efficiency, is generally not feasible by examining the reaction time course only. Instead, it is necessary to utilise the concentration *vs* time data available from experimental kinetic analysis of a particular replicator and undertake full kinetic fitting, taking into account all of the reaction and recognition channels available to the system, in order to obtain these parameters. Only comprehensive kinetic fitting can determine the contribution of the various reaction pathways towards product formation.

1.4.5 Synthetic replicating systems in practice

The theoretical requirements for the establishment of self-replicating systems were complemented by reports of abiotic experimental systems, expressing the capacity to template their own synthesis. These synthetic replicating systems, ranging from simple replicators operating in isolation to a variety of increasingly complex interconnected systems, can facilitate our understanding of the molecular origins of life.

1.4.5.1 Oligonucleotide-based replicators

Following extensive work on template-directed synthesis[75–78] of oligonucleotides by Orgel and co-workers, the von Kiedrowski laboratory reported[79] in 1986 the first example of non-enzymatic self-replication in a model chemical system based on a oligonucleotide strand with a palindromic sequence (**Scheme 1.1**). In the replication cycle, a trinucleotide **CCG** (protected at the 5' end), activated using carbodiimide EDCI *in situ*, was coupled to trinucleotide **CGG** (protected at the 3' end), producing a hexanucleotide template **CCGCGG**. The complementarity of the template sequence, now protected at both the 3' and the 5' end, to the building blocks **CCG** and **CGG** enabled the association of the template with the trinucleotide precursors *via* hydrogen-bonding-mediated recognition between the Watson-Crick base pairs, thus facilitating the formation of further template molecules.



Scheme 1.1 A self-replicating hexanucleotide with a palindromic **CCGCGG** sequence capable of catalysing its own formation by associating of template sequence with two smaller trinucleotides **CCG** and **CGG** in a catalytically active ternary complex, as reported by von Kiedrowski in 1986. Scheme adapted from Ref. 79.

The autocatalytic reaction order was shown to be low ($p = 0.48$), suffering from significant product inhibition. The ability of the hexanucleotide to replicate was improved when the sequence was redesigned[80] to contain an amine functional moiety on the cytosine in **CGG** trinucleotide instead of a hydroxyl group. While this phosphoamidate linkage based replicator maintained its parabolic growth, the reaction profile exhibited a sigmoidal reaction profile and good autocatalytic efficiency ($\varepsilon = 420$). In further work, Kiedrowski and co-workers have, in an attempt to overcome the product inhibition that limited the replication efficiency of the previous designs, successfully employed solid support-immobilised DNA strands, in order to achieve[81] exponential replication.

Building on the design of a single replicating system, the Kiedrowski laboratory has extended[82,83] their oligonucleotide-based replicator into a multicyclic system (**Figure 1.10**), capable of both autocatalysis and crosscatalysis. The new network was composed of four DNA-based trinucleotides: **Ap** (**CCG**), **nB** (**CGG**), **nA** (**CCG**) and **Bp** (**CGG**). Trinucleotides **Ap** and **Bp** were equipped with a reactive electrophilic

phosphate group (denoted by **p**) at the 3' end, while the 5' prime end of these two components was protected in order to prevent any self-condensation reactions from occurring. Trinucleotides **nB** and **nA** incorporated a free nucleophilic amine (marked as **n**) at the 5' end. As for **Ap** and **Bp**, their 3' phosphate group was protected, in order to prevent any unwanted ligation reactions.

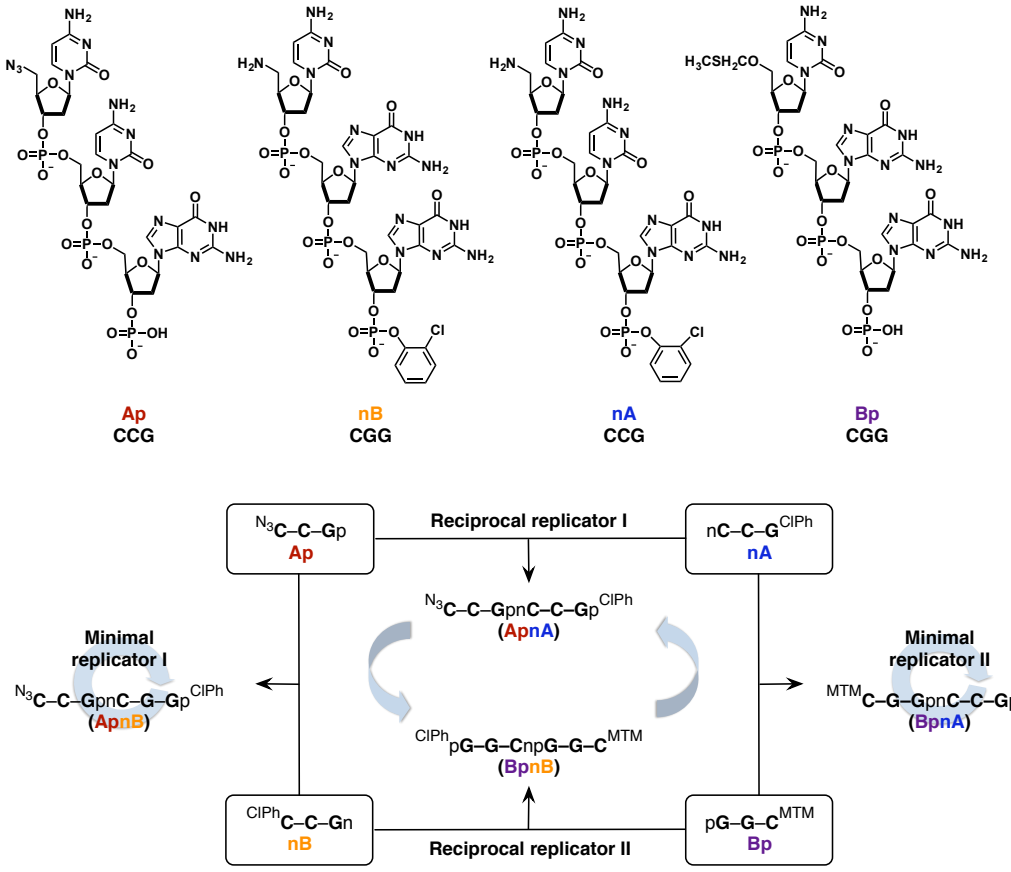


Figure 1.10 Multicyclic replicating network composed of four trinucleotides: **Ap** (**CCG**), **nB** (**CGG**), **nA** (**CCG**) and **Bp** (**CGG**). The components can react to form two self-replicating templates (**ApnB** and **BpnA**) and two reciprocally replicating templates (**ApnA** and **BpnB**). The letters **n** and **p** correspond to the presence of a free amine group at the 5' end and phosphate group at the 3' end, respectively. Figure adapted from Ref. 84.

When reacted together, these four trinucleotide building blocks can form four templates. Condensation of **Ap** and **nB** results in the formation of a self-complementary autocatalytic template **ApnB** (**CCGCGG**). Similarly, reaction of **nA** with **Bp** yields a self-replicating template **BpnA** with sequence **CGGCCG**. The hexanucleotide templates, produced by the reaction of **Ap** with **nA** (**CCGCCG**) and **Bp** with **nB** (**GGCGGC**) are capable of reciprocal replication only. In this multicyclic system, formation of template from the components assembled in a ternary complex proceeded by an attack of the 5' amine on the 3' phosphate, at a comparable rate for all templates. Recognition in all four ternary complexes was mediated by Watson-Crick base pairing, and as a

result of these similarities, the four template products were nearly evenly distributed in the uninstructed experiment. Addition of preformed reciprocal template resulted in an increase in the concentration of the complementary strand. When the replicating network was instructed with template of either of the two self-replicating sequences, up-regulation (increase in the concentration) of both self-replicating pathways was observed. Interestingly, the second self-replicator was present at an increased concentration as well. The fact that instructing the system with one self-replicating template resulted in increased formation of the second self-replicators is a system-level property, stemming from the fast consumption of the components necessary for the reciprocal cycles to operate efficiently.

In 2002, Paul and Joyce reported[85] the first example of an RNA-based self-replicating system, exploiting[86] an R3C RNA ligase ribozyme. The study employed an adapted version of an R3C ligase ribozyme, capable of catalysing the formation of a 3',5'-phosphodiester bond between two individual RNA molecules. The RNA ribozyme template **T** was designed to be capable of ligating two RNA subunits **A** and **B** (**Figure 1.11**) *via* a ternary complex, producing an exact copy of itself. The reaction was mediated by a nucleophilic attack of the 3'-hydroxy group of **A** on the α-phosphate of the 5'-pppG of the subunit **B**, affording a template duplex [**T**·**T**], with a two-fold centre of symmetry.

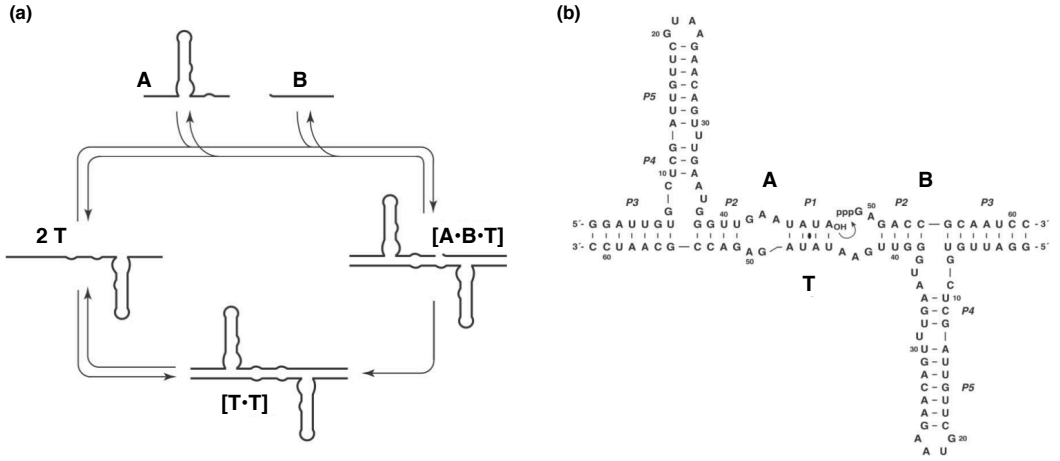


Figure 1.11 **(a)** Self-replication of an RNA ribozyme follows the minimal replication-model of self-replication. Template **T** can assemble with **A** and **B** into a ternary complex, facilitating the formation of new template. **(b)** Structural details of the self-replicating ligase ribozyme, with two-fold symmetry in the middle. Figure taken from Ref. 85. Copyright 2002 National Academy of Sciences, USA.

When the formation of the RNA ribozyme was examined in the presence of preformed template, the system showed an increase in the initial rate of template formation, with an autocatalytic order $p = 1$. Kinetic fitting revealed that two processes contribute to the formation of template **T**: reaction of preformed complex [**B**·**T**] with component **A** and slow bimolecular reaction of **A** and **B**. The enhanced reaction rate in the template-

instructed experiment, however, was found to be only temporary, indicating that while the ternary complex contributes to the formation of template, a competing process in the system prevents the replication cycle from operating efficiently. The addition of preformed template to the reaction mixture accelerated template formation only for a limited amount of time, and the authors suggested that the lower than expected efficiency of the designed RNA system stems from the similarities in the nucleotide sequences of components **A** and **B**, resulting in inhibition through the formation of an inactive binary complex [**A**·**B**], the binding of which cannot be disrupted by the added preformed template. The authors found that the negative effect of strong [**A**·**B**] complex on replication can be circumvented by premixing **T** and **B**, prior to addition of **A**. Alternatively, addition of excess **A** to the reaction mixture achieved a similar effect.

Joyce and Kim have expanded[87] the R3C ligase self-replicating system into a reciprocal replicating network (**Figure 1.12**), where two ribozymes, **T** and **T'**, template the formation of each other from the corresponding subunits:**A**, **B**, **A'** and **B'**. While the ribozymes contain complementary recognition and catalytic elements, the two sequences are not identical.

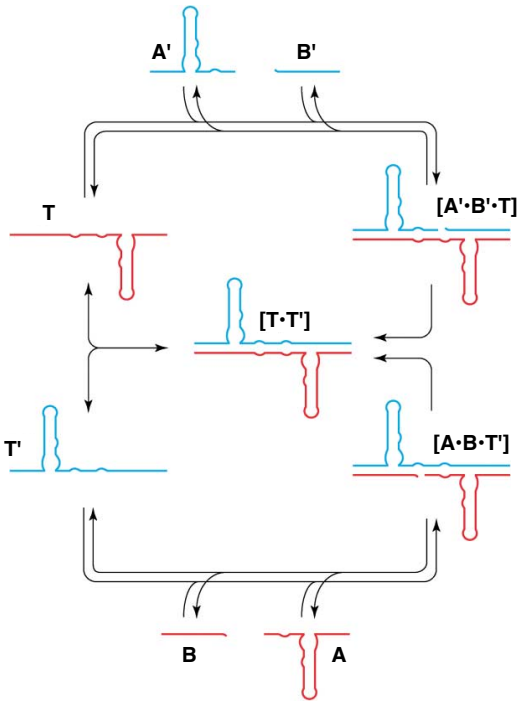


Figure 1.12 Crosscatalytic replication network of two R3C ligase ribozymes. Ribozyme **T'** (blue) assembled from components **A'** and **B'** catalyses the ligation reaction between **A** and **B** to form RNA ribozyme **T** (red). Similarly, the product **T** mediates the formation of **T'** from **A'** and **B'**. Both reactions, mediated by reciprocal template proceed through catalytically-active ternary complexes [**A'**·**B'**·**T**] and [**A**·**B**·**T'**]. Figure adapted from Ref. 87 with permission from Elsevier.

The loss of self-complementarity prevents the association of the substrate subunits, which hindered replication in the self-replicating system described above. As previously, each ribozyme catalyses its synthesis *via* the assembly of reaction components in a ternary catalytically-active complex ([**A'**·**B'**·**T**] and [**A**·**B**·**T'**]). Kinetic analyses of the individual crosscatalytic pathways and the reciprocal system as a whole, showed that both templates catalyse the formation of the reciprocal replicator and result in increased formation of the ribozyme itself as well.

1.4.5.2 Peptide-based replicators

In the synthetic oligonucleotide-based synthetic replicators discussed in the previous section, transfer of information depends on a fairly specific, well-defined pattern of molecular recognition elements. Namely, recognition is afforded by hydrogen-bonding interactions between donor and acceptor elements, encoded within each oligonucleotide sequence. Formation of a self-replicating peptide, on the other hand, necessitates that the peptide template be able to associate with smaller peptide fragments in some form of a catalytically-active complex. Oligopeptides possess an extremely rich structural lexicon, arising from the increased number of building blocks used—20 amino acids compared to four nucleotides in replicating systems based on DNA. Inter peptide recognition is determined not only by the primary sequence of amino acids in each sequence, but also by the secondary and tertiary structures governed by the interactions between those amino acids, increasing the potential challenge of designing a self-replicating peptide.

The first experimental demonstration of peptide replication was reported[88] in 1996 by Ghadiri and co-workers. The design of the 32-residue self-replicating α-helical peptide (**Figure 1.13**) was inspired[89] by the leucine zipper domain of the yeast transcription factor GCN4. This peptide replicator design exploited a simple protein folding motif—an α-helical coiled-coil—distinguished by peptide sequences composed of heptad repeats, resulting in two coiled-coils wrapped around each other with a slightly left-handed, superhelical twist. The sequence of the reported peptide replicator (**Figure 1.13**) implements six substitutions relative to the wild type GCN4. Of particular interest is the substitution of a neutral, hydrophilic asparagine residue (position 16 in the sequence), located within the core hydrophobic region, with a hydrophobic valine residue, which enabled[90] equilibration between a dimeric and trimeric coiled-coil structure.

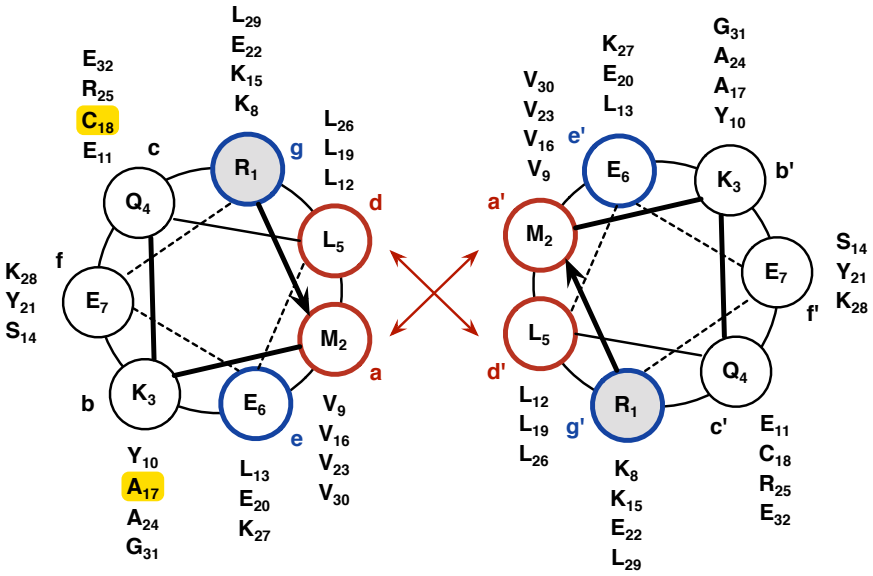


Figure 1.13 Design of an α-helical, coiled-coil peptide capable of self-replication, reported by Ghadiri and co-workers in 1996, featuring a heptad repeat (abcdefg). Recognition between peptides and their assembly into complexes is mediated by the recognition between the hydrophobic residues at position a and d (red), and electrostatic interactions between residues at position e and g (blue). Residues b, c and f (black) are exposed to the solvent and do not contribute to the recognition. Position of the two residues required for native chemical ligation, alanine (activated as thiobenzyl ester) and cysteine, is highlighted in yellow. Figure adapted from Ref. 88.

Monomeric coiled-coil peptides are generally present as random coils in aqueous solutions. However, these peptides can adopt a completely α-helical structure, providing that a suitable template framework for directing their assembly is present. As with other minimal replicating systems, an autocatalytic peptide system built from two smaller complementary peptides, each equipped with a reactive group, needs to assemble on a peptide sequence that positions these fragments in an orientation that facilitates their reaction. In a situation where these fragments are the constituent parts of a longer template sequence, the product formed by their reaction constitutes an identical copy—opening up the possibility of self-replication (**Figure 1.14a**). The Ghadiri peptide described here is capable of forming both dimeric and trimeric assemblies and the authors proposed that both **T** and [**T**·**T**] could potentially serve as a template in the autocatalytic pathway.

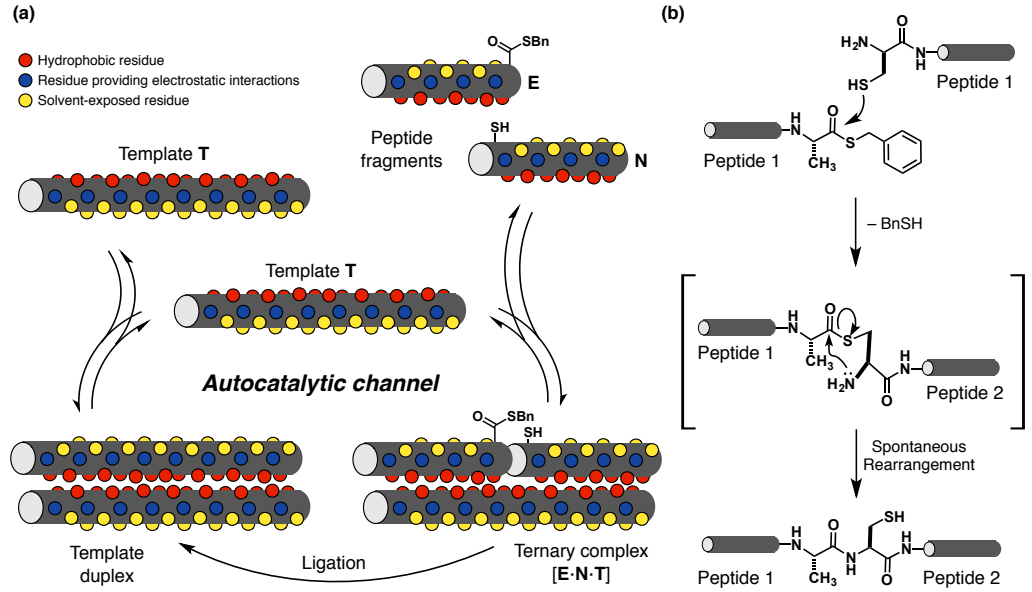


Figure 1.14 **(a)** Schematic representation of peptide self-replication, where an electrophilic component **E** and nucleophilic fragment **N** react through a bimolecular reaction to give template **T**. The peptide template can assemble the respective fragments in to a ternary, catalytically-active complex [**E**·**N**·**T**] which facilitates the ligation between the two reactive sites on **E** and **N**. The ligation produces a template duplex, the dissociation of which is necessary in order for efficient self-replication. **(b)** Mechanism of the native chemical ligation reaction that leads to the formation of template. The ligation occurs between the N-terminal preactivated thioester on **E**, and the C-terminal cysteine located on the nucleophilic fragment **N**. Figure adapted from Ref. 88.

The recognition mediating the template-directed reaction in the Ghadiri peptide was afforded through the interactions of complementary hydrophobic and electrostatic peptide surfaces. Specifically, residues at position a and d (**Figure 1.13**, red) within the peptide sequence drive the inter-helical recognition through hydrophobic interactions, playing a pivotal role in determining the stability and orientation of coiled-coil peptides. Residues in position e and g (**Figure 1.13**, blue) within the heptad repeat are responsible for driving the intra-component recognition through electrostatic interactions. Residues b, c and f (**Figure 1.13**, black), on the other hand, are located on the solvent exposed surface, and, thus, do not contribute to the recognition. The ligation site (**Figure 1.13**, yellow) is positioned on the solvent-exposed surface, in order to avoid potential interference with the hydrophobic core responsible for recognition. The peptide coupling strategy exploited by Ghadiri and co-workers employed a thioester-promoted native peptide ligation (**Figure 1.14b**), first described[91] by Kent and co-workers in 1994. Peptide template **T** is formed through the reaction of an N-terminal, 17-residue electrophilic fragment **E**, activated as a thiobenzyl ester and a 15-residue C-terminal nucleophilic fragment **N**, bearing a free cysteine residue. The native ligation reaction proceeds through an intermediate thioester (**Figure 1.14b**), which undergoes

intramolecular rearrangement to produce the final, more thermodynamically stable amide bond at the ligation site.

The ability of the designed coiled-coil peptide to self-replicate was established unambiguously through template-instructed kinetic experiments, where the ligation reaction was examined in the presence of increasing quantities of preformed peptide template **T**, added at $t = 0$. The replication profile displayed parabolic growth, where the initial rate of ligation correlated with the square root of the concentration of the initial template added, suggesting that replication is limited by product inhibition despite the relatively high autocatalytic efficiency ($\varepsilon = 500$). Through kinetic fitting and simulation of the experimental data, the authors were able to establish the reaction order p as 0.63. This reaction order is higher than the reaction order observed generally in nucleotide-based self-replicating systems, and may potentially arise from catalysis through quaternary complexes, mediated by template duplex [**T**·**T**].

By exploring conservative substitutions of the residues at the key positions within the sequence of the peptide, a and d, responsible for molecular recognition, the authors were able to demonstrate that the efficiency of the replication mechanism is extremely sensitive to the identity of the residues within the peptide sequence. In particular, the authors analysed two conservative mutations, where an alanine residue was substituted for valine (residue 9) within the heptad repeat, and instead of leucine at position d (residue 26). Despite the conservative nature of the mutations in the residues responsible for the hydrophobic interactions, these altered peptides showed no significant template-assisted catalytic activity.

Reliance of the self-replication mechanism on molecular recognition was demonstrated through two carefully designed control experiments. The reaction between **E** and **N** was examined in the presence of guanidinium hydrochloride, a chaotropic reagent, which exerts a destabilising effect on complexes in the system, hindering the ability of the system to partake in recognition-mediated reactions. This experiment showed that the concentration-time profile in the presence of such chaotropic reagent closely mirrors the reaction profile determined for the background, uncatalysed reaction, with a concomitant loss of the sigmoidal reaction profile. Furthermore, no enhancement in the rate of formation of **T** was observed in the presence of added preformed template under these conditions. The second control experiment was specifically designed to probe whether the reactions of binary complexes with the individual smaller fragments, *i.e.* [**T**·**E**] with **N** and [**T**·**N**] with **E**, contribute to the production of peptide **T**. These experiments employed "crippled" peptide sequences, each containing a single mutation within the hydrophobic recognition-mediating core of both peptide fragments, namely substitution of a glutamic acid residue in place of valine (position 9) and leucine (posi-

tion 26). Kinetic analyses confirmed that the addition of the mutated templates (formed by the reaction of a "crippled" and native fragment), capable of associating with **E** or **N** in to binary complexes only, afforded no enhancement in the rate of formation of the native peptide. Taken together, the authors were able to establish unambiguously that a recognition-mediated enzyme-free peptide replication is possible in systems exploiting the coiled-coil structural motif.

The initial report of a self-replicating peptide by Ghadiri and co-workers was soon followed by several other reports of peptide replicators, exploiting similar design principles. Utilising the coiled-coil helical peptide platform, Chmielewski and co-workers have reported[92,93] two examples of peptide systems that could be modulated through environmental control, and, thus, allowing self-replication to be turned on and off selectively. Specifically, the Chmielewski laboratory showed that self-replication in a peptide system can be tuned by pH,[92] as well as through ionic control.[93] The concept of environmental control can be illustrated on the pH modulated replicator **E1E2** (**Figure 1.15**), for example, formed by reaction of two subunits, **E1** and **E2**, incorporating two glutamate residues (at position e and g), protonated under acidic conditions. At physiological pH, however, these glutamate residues are negatively charged, resulting in destabilisation of the coiled-coil assembly. The random coil conformation adopted by the peptide at neutral pH is incapable of supporting self-replication. Therefore, the peptide can replicate successfully only at low pH (≤ 4.0), when the glutamate residues, essential for recognition, are protonated. Satisfyingly, the authors were able to establish through template-instructed experiments that under these conditions, the catalytic efficiency is similar to that observed in the Ghadiri system (ε = 900).

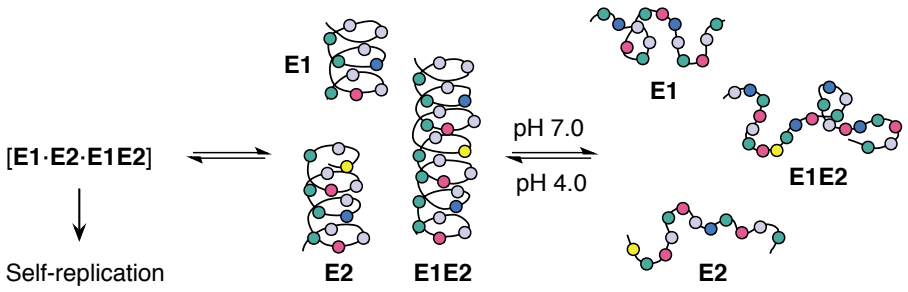


Figure 1.15 Design of a self-replicating peptide **E1E2** modulated by pH, as described by Chmielewski and co-authors. The recognition-mediated reaction processes in the system, and, thus also the formation of catalytically-active complexes, are only effective at low pH (pH = 4), conditions at which the two glutamate residues are protonated. Figure adapted from Ref. 92.

As often observed with synthetic replicating systems based oligonucleotides, the Ghadiri and Chmielewski replicating peptides, while capable of templating their own formation, suffered from significant product inhibition. In the coming years, remarkable results in overcoming product inhibition in peptide-based replicating systems have been achieved by the Chmielewski laboratory in particular. Chmielewski and co-workers have explored two strategies for increasing the efficiency of replication. In 2002, Issac and Chmielewski exploited the findings reported[94,95] in the literature showing that the stability of coiled-coil assemblies can be modulated by altering the length of the peptide sequence. The authors adapted[96] the sequence of the **E1E2** peptide described previously, shortening it by one heptad repeat. The modification resulted in a self-replicating peptide with a dramatically icnreased catalytic efficiency ($\varepsilon = 500000$) and reaction order $p = 0.91$. In an alternative strategy, Li and Chmielewski achieved[97] exponential replication by introducing a proline residue at a strategic position within the peptide sequence, thus destabilising the product duplex. The replication efficiency of the proline-containing replicator increased to $\varepsilon = 320000$, with a reaction order similar to that observed in the shorter-sequence peptide replicator.

Both Ghadiri and Chmielewski laboratories have been successful in developing these examples of individual self-replicating peptides into more complex networks where multiple catalytic pathways and replicators operate simultaneously. Chmielewski and co-workers have combined[98] the two environmentally-responsive self-replicating peptides into a single system, where both auto- and crosscatalytic cycles are active. The expanded peptide network was assembled from four shorter peptide fragments, **E1**, **E2**, **K1** and **K2**, which permitted formation of the two native peptide templates, **E1E2** and **K1K2**, and two recombinant proteins, **E1K2** and **K1E2**. These mixed templates are capable of self-associating *via* anti-parallel coiled-coils and capable also of associating with each other *via* formation of parallel coiled-coils. Kinetic analyses of the various reaction pathways showed that under neutral pH conditions, the recombinant **E1K2** template is produced most rapidly. Despite this preference of the system towards production of **E1K2**, the authors were able to selectively amplify **E1E2** product by decreasing the pH of the reaction environment to 4. Similarly, the authors were able to direct the network towards enhanced production of **K1K2**, by undertaking the reaction under high salt conditions (at neutral pH). Using this framework, Chmielewski and co-authors demonstrated successfully that the production of a particular peptide replicator can be amplified selectively from a mixture of reactive components by modulation of the reaction environment, such as the pH or salt concentration, providing support for the potential role of proteins in the emergence of life.

The first multicyclic network explored by Ghadiri laboratory reported[99] an example of a symbiotic, mutually auto- and crosscatalytic peptide network, where two replicators were capable of templating their own formation as well as the formation of each other. In fact, as a result of the higher catalytic efficiency of the crosscatalytic pathways relative to the autocatalytic ones, both replicators were able to coexist and enhance formation of each other within the network. In further work, Ghadiri and co-workers exploited[100] the sensitivity of the coiled-coil peptide replicator framework to changes in the residues required for recognition in designing a dynamic peptide network capable of error-correction (**Figure 1.16**).

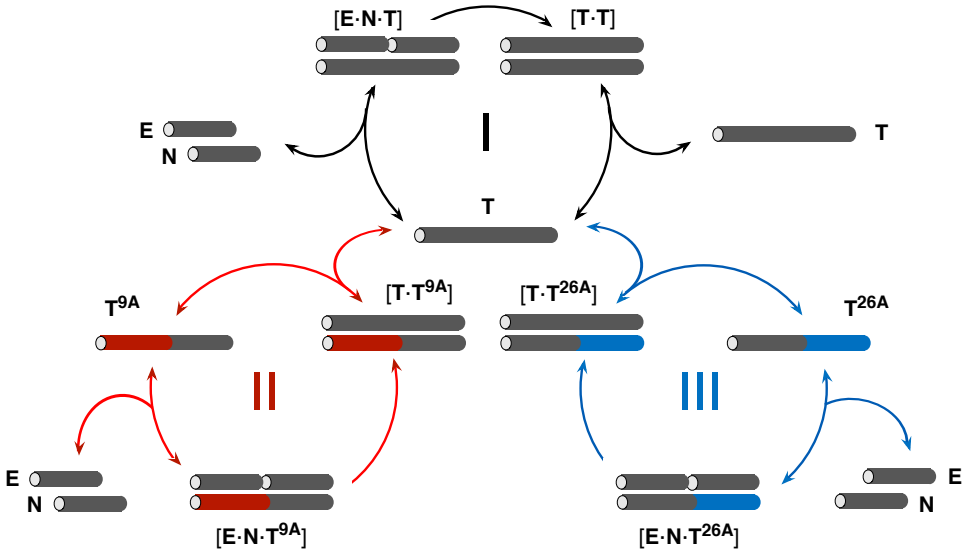


Figure 1.16 Schematic representation of the recognition-mediated catalytic pathways active in an error-correcting, autocratic peptide network. A mixture of peptide fragments **E** and **N**, and their single-alanine mutants, $\mathbf{E^{9A}}$ and $\mathbf{N^{26A}}$, (simulating spontaneous generation of errors) results in a wild type template **T** (grey cylinder) and single mutation containing templates $\mathbf{T^{9A}}$ (red cylinder) and $\mathbf{T^{26A}}$ (blue cylinder). The self-organised network amplifies the template **T** selectively by subjugation of the mutant templates for the production of **T**. The double mutant $\mathbf{T^{9A/26A}}$ is not shown as it was determined to be catalytically inactive. Figure adapted from Ref. 100.

The authors achieved selective amplification of a single peptide replicator within this simultaneously auto- and crosscatalytic system by recruitment of the mutant peptides for the synthesis of the wild type peptide, **T** (**Figure 1.16**, grey). Slow spontaneous generation of errors/mutants, as observed traditionally in biological systems over time, was simulated by formation of structurally-related mutant peptides through bimolecular reaction of smaller fragments incorporating mutations. In addition to the native electrophilic and nucleophilic fragments, **E** and **N** (**Figure 1.16**, grey), the network included their single alanine mutants, $\mathbf{E^{9A}}$ (**Figure 1.16**, red) and $\mathbf{N^{26A}}$ (**Figure 1.16**, blue). Reaction of these fragments afforded four different peptide templates, the native

T, $\mathbf{T^{9A}}$ and $\mathbf{T^{26A}}$ with a single mutated residue and double mutant $\mathbf{T^{9A/26A}}$. Under neutral conditions, the reaction system showed strong preference for the formation of the mutation-free species **T**. The double mutant $\mathbf{T^{9A/26A}}$ was shown to be completely catalytically-inactive, whereas the two templates incorporating a single mutation were capable of crosscatalytic activity only, directed towards the enhanced formation of the native replicator **T**. Interestingly, the error-free template **T** was found to be a selfish autocatalyst, which means that the autocatalytic cycle producing **T** worked in concert with the two crosscatalytic pathways to achieve selective production of **T**. Within this peptide network, the authors have demonstrated an example of a peptide network capable of exhibiting two complex phenomena simultaneously, error-correction and sequence-specific replication, with potential significance in genotype stabilisation of self-replicating molecules.

The world as we know it is homochiral, yet, the origins of this biological homochirality have yet to be established and are a source[101–104] of ongoing debate. Exploring the possible role of peptide replicators in this process, Ghadiri and co-workers have designed[105] a replicating peptide network capable of stereospecific replication (**Figure 1.17**).

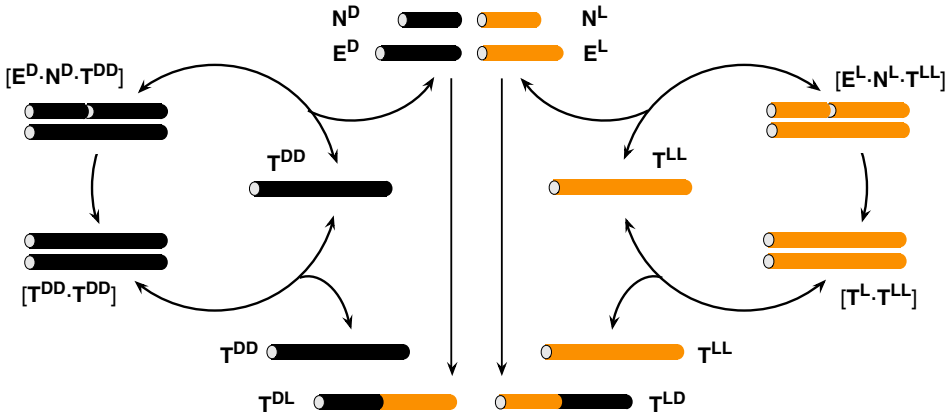


Figure 1.17 Schematic representation of stereospecific peptide replicators. The electrophilic fragments, $\mathbf{E^L}$ and $\mathbf{E^D}$, and nucleophilic fragments, $\mathbf{N^L}$ and $\mathbf{N^D}$, combine to form four templates. The homochiral templates $\mathbf{T^{LL}}$ and $\mathbf{T^{DD}}$ are capable of autocatalysis, while the heterochiral templates can only be formed through uncatalysed bimolecular reactions. Black and orange cylinders represent peptide regions comprised of D- and L-amino acids, respectively. Figure adapted from Ref. 105.

This work extended their original peptide replicator **T** to a system composed of two electrophilic fragments, $\mathbf{E^L}$ and $\mathbf{E^D}$, and two nucleophilic components, $\mathbf{N^L}$ and $\mathbf{N^D}$. Reaction of fragments with the same stereochemistry produced homochiral templates $\mathbf{T^{LL}}$ (**Figure 1.17**, orange) and $\mathbf{T^{DD}}$ (**Figure 1.17**, black), whereas reaction of mixed fragments created heterochiral $\mathbf{T^{LD}}$ and $\mathbf{T^{DL}}$. Examination of the various reactions

between the nucleophilic and electrophilic components showed that the homochiral products are formed efficiently and preferentially.

The heterochiral peptide templates, $\mathbf{T^{LD}}$ and $\mathbf{T^{DL}}$ were shown to form through template-independent pathways only, and the authors suggest that this observation stems from the diminished ability of these two templates to form coiled-coil helical assemblies. Detailed kinetic analysis and template-instructed experiments revealed that $\mathbf{T^{LL}}$ is capable of stereospecific self-replication, insensitive to addition of the homochiral $\mathbf{T^{DD}}$ or the heterochiral templates. The chiroselective system exhibited strong sensitivity to mutations in even a single amino acid residue, which contributed to amplification of a single homochiral template, once formed. While the authors have, for the first time, demonstrated the feasibility of chiroselective replication, the ability of the homochiral template $\mathbf{T^{DD}}$ to self-replicate was not discussed in the study and, therefore, no conclusions can be drawn as to whether stereoselective replication is present in both homochiral replicators. Nevertheless, these results suggest that a peptide biopolymer could have played a role in the origin of biological homochirality.

The design of peptide replicators continued to evolve, exploring larger and more interconnected systems. In 2004, Ghadiri and co-workers described[106] a bottom-up approach to designing a peptide network composed of 81 structurally-similar 32-residue coiled-coil peptides (**Figure 1.18**). Analysis of the numerous peptides relied initially on computational prediction methods in order to estimate the relative stability of all substrate-template complexes available to the system. The authors analysed the stability information of each complex in order to predict the auto- (**Figure 1.18**, red arrows) and crosscatalytic pathways (**Figure 1.18**, black arrows) and as well as the network topology.

The employed design principles were tested experimentally on a smaller, 9-node subsystem (**Figure 1.18**, dark grey) within the network, which showed a good agreement with the estimations predicted through graph analysis. Ghadiri and co-workers also demonstrated that the efficiency of certain network pathways can be selectively modulated by employing various chemical triggers, *i.e.* different instructing templates. The smaller 9-peptide system was built from a single nucleophilic component **N** and nine different electrophilic peptide fragments, $\mathbf{E^1}$ to $\mathbf{E^9}$. The authors exploited substitutions at four key electrophilic residues, located at positions e and g within the peptide sequence (**Figure 1.18**, represented by the four letter code), in order to design peptides with varied ability to form aggregates, and, thus, also different catalytic efficiencies.

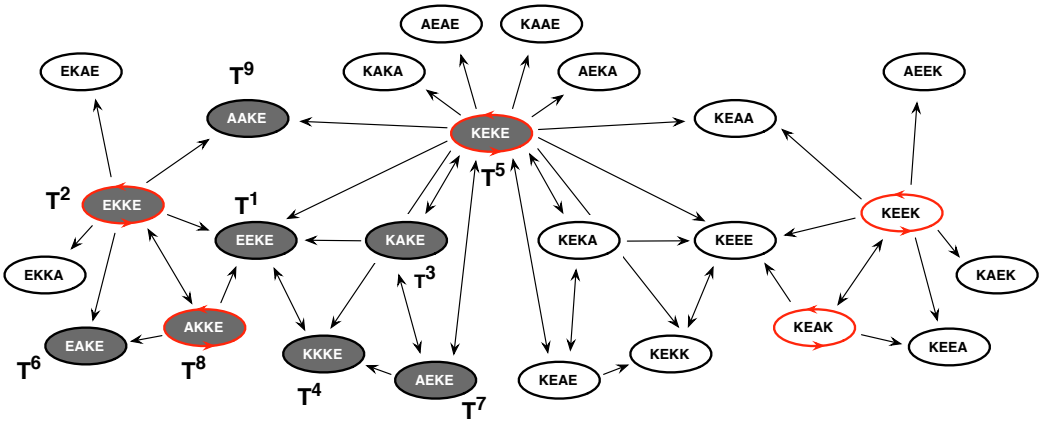


Figure 1.18 Graph representation of a directed self-organised peptide network established from an array of 81 structurally-similar coiled-coil sequences (32 residues each) in silico. The network is comprised of 25 nodes (molecular species) and joined by 53 edges (arrows), representing auto- (red circular arrows) and crosscatalytic (black arrows, where the arrow direction indicates catalytic relationship) processes. Sub-network of nine nodes, $\mathbf{T^1}$ to $\mathbf{T^9}$, highlighted in dark grey, was investigated experimentally, showing a good agreement with the predicted results. The four letters within each node (template) represent the identity of the amino acids within the peptide sequence at residues 8, 13, 15 and 20 (positions e and g). Figure adapted from Ref. 106.

Theoretical analysis predicted that twenty crosscatalytic pathways and three autocatalytic pathways are plausible in this network. Experimental analysis of a mixture containing all of the electrophilic fragments with a sub-stoichiometric amount of **N** revealed that all nine possible products are formed, with $\mathbf{T^1}$, $\mathbf{T^2}$, $\mathbf{T^4}$, $\mathbf{T^7}$ and $\mathbf{T^8}$ reaching the highest concentrations. Undertaking a comprehensive kinetic analyses of the individual reaction pathways as well as the network as a whole showed that the rate of formation of the examined templates was noticeably different in isolation relative to the rate when the full network was examined. These differences highlight the potential of system-level properties manifesting in networks of interconnected components that are not observed when the components are examined in isolation.

Ghadiri and Ashkenasy examined[107] a sub-network of this smaller 9-node system, demonstrating its capacity to perform basic Boolean logic functions, such as OR and NOR (neither X nor Y), when instructed with chemical input (preformed templates)—control operations not unlike those observed in complex biological systems. The system (**Figure 1.19a**) is comprised from five nodes (peptide templates $\mathbf{T^1}$, $\mathbf{T^3}$, $\mathbf{T^4}$, $\mathbf{T^5}$ and $\mathbf{T^7}$) that are interconnected through 15 edges, representing auto- and crosscatalytic pathways.

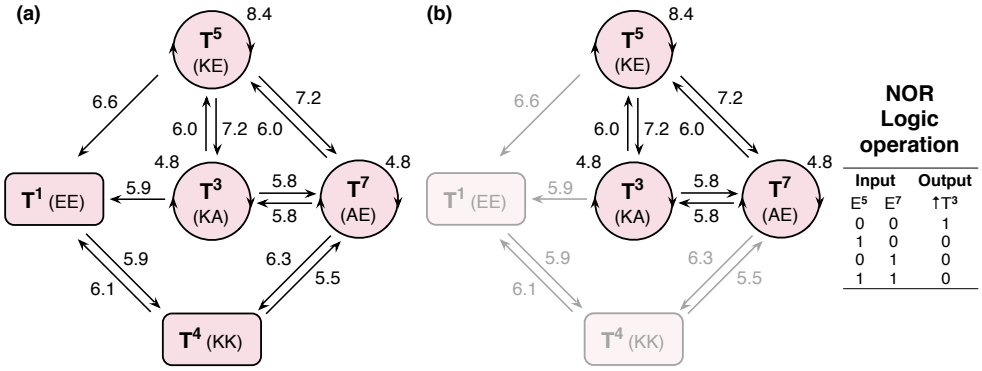


Figure 1.19 **(a)** A directed, weighted peptide network capable of performing logic operations. Five peptide templates are interconnected through 15 auto- and crosscatalytic pathways. **(b)** Sub-network of three templates, $\mathbf{T^3}$, $\mathbf{T^5}$ and $\mathbf{T^7}$ expresses NOR logic operation, where self-replication of $\mathbf{T^3}$ proceeds efficiently only in the absence of both electrophilic fragments $\mathbf{E^5}$ and $\mathbf{E^7}$ Figure adapted from Ref. 107.

As an example, the NOR logic operation (**Figure 1.19b**) is expressed by the auto-catalytic formation of $\mathbf{T^3}$. In this Boolean function, template $\mathbf{T^3}$ is formed efficiently only in the absence of both the $\mathbf{E^5}$ and $\mathbf{E^7}$ fragments. If either of these fragments is present, however, formation of $\mathbf{T^3}$ is diminished as any quantity produced is utilised primarily as a crosscatalytic template for the formation of $\mathbf{T^5}$ and/or $\mathbf{T^7}$.

An alternative strategy, exploiting light as a trigger for exerting control over logic operations in peptide replication networks, was introduced[108] by Ashkenasy and co-workers. The experimental design was based on a dimeric coiled-coil assembly exhibiting high sequence selectivity. The peptide template was designed to contain a photocleavable moiety (**Figure 1.20**, yellow star), 6-nitroveratryloxycarbonyl (**Nv**) bound to a lysine residue in position g, responsible for electrostatic interactions. The addition of this caging element afforded a peptide template that has a significantly reduced propensity for dimerisation and association with smaller peptide fragments, **N** (**Figure 1.20**, blue) and **E** (**Figure 1.20**, green), when present. Exposure to light resulted in efficient removal of the **Nv** group, thereby re-establishing the ability of the template to replicate through the template-mediated pathway.

Ashkenasy and co-workers demonstrated the that amount of catalytically-active template within the mixture can be modulated by the length of exposure of the system to light, thereby providing a direct control over the rate of replication. This concept of light-induced replication was examined further utilising two nucleophilic fragments, $\mathbf{N^{aa}}$ and $\mathbf{N^{z}}$, competing for the shared electrophilic peptide. These two additional fragments incorporated a leucine to alanine mutation within the hydrophobic recognition core and an additional caging moiety, respectively. Reaction of a mixture containing all of the building blocks was studied under four different conditions, altering the presence

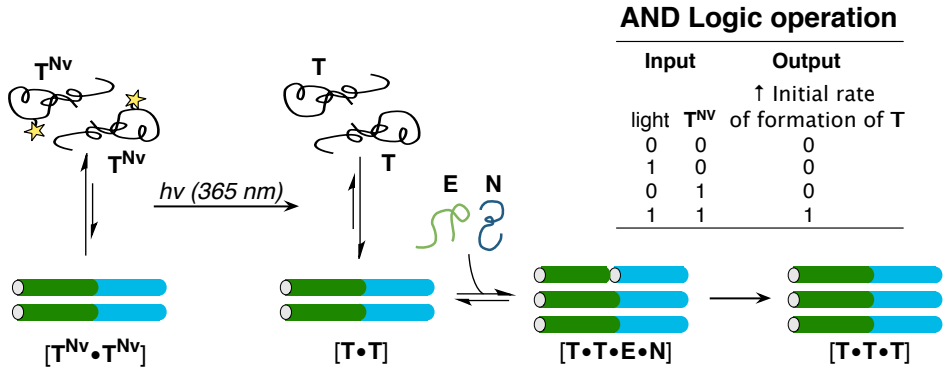
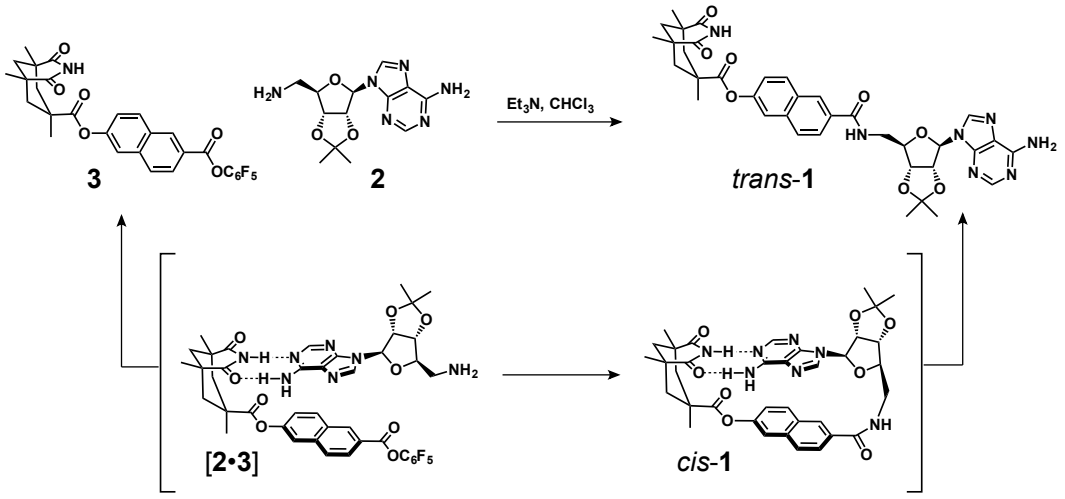


Figure 1.20 Replicating peptide system controlled by light. In the absence of light (left), the peptide template $\mathbf{T}^{\mathbf{Nv}}$ exists as a random coil, incapable of dimerisation. Exposure to monochromatic UV light cleaves the caging moiety (yellow star, position g), producing an active template **T**. The activated template forms a dimeric coiled-coil [**T**·**T**] that is capable of associating with smaller fragments, **E** (green) and **N** (blue) in a quaternary complex [**T**·**T**·**E**·**N**], catalysing formation of further template molecules. Initial rate of formation of **T** in the presence of light and $\mathbf{T}^{\mathbf{Nv}}$ as the stimulus is governed by the AND logic operation. Adapted from Ref. 108.

of $\mathbf{T}^{\mathbf{Nv}}$ and light. The outcome showed that formation of template **T** is governed by the AND logic operation (**Figure 1.20**), where presence of both light and preformed $\mathbf{T}^{\mathbf{Nv}}$ is required in order for **T** to form selectively and efficiently from the mixture of components.

1.4.5.3 Small molecule-based replicators

A variety of molecules capable of self-replication inspired by oligomers found in nature have been presented thus far. Building on these examples, the Rebek laboratory demonstrated[109] the first example of a small molecule-based self-replicator, in 1990. The system exploited a Et_3N-catalysed formation of an amide bond as the strategy for template formation. Specifically, template **1** was formed by the reaction of an adenine derivative **2** and an imide of Kemp's triacid[110] **3** (**Scheme 1.2**). Rebek's design, however, did not exhibit a sigmoidal reaction profile. This was attributed to the binary reactive complex [**2**·**3**] pathway being more efficient than the ternary complex [**2**·**3**·**1**]. Reaction within the complex [**2**·**3**] produced *cis*-**1** template, which further isomerised to give template *trans*-**1**.



Scheme 1.2 Rebek's self-replicating system. Components **2** and **3** can react to form template **1**. The reaction can also proceed through a [**2**·**3**] binary complex, affording product *cis*-**1**, which can isomerise to give the more stable template *trans*-**1**. Scheme adapted from Ref. 109.

Rebek *et al*. were able to demonstrate that self-replication is recognition-mediated when they observed a drop in reaction rate when **3** was reacted with a recognition-disabled, *N*-methylated version of **2**, or in the presence of a competitive inhibitor: 2,6-bis(acylamino)pyridine. In 1994, Menger published[111] a study demonstrating that the formation of template **1** can be catalysed by addition of simple amides, thus raising doubts as to the self-replicating nature of Rebek's system. Following a prolonged debate, the Reinhoudt laboratory provided[112] the evidence that finally allowed the argument between Rebek and Menger to be settled in 1996. Through a full kinetic analysis of the replicating system, they were able to identify five different pathways through which the replication proceeds (**Table 1.1**).

Table 1.1 Five replication pathways identified[112] by Reinhoudt and co-workers for the Rebek self-replicating system. Recognition-mediated complex formation is denoted by square brackets and reaction between components by +.

Pathway				
I	**II**	**III**	**IV**	**V**
2 + **3**	[**2**•**3**]	[**2**•**3**•**1**]	[**3**•**1**] + **2**	**3** + [**2**•**1**]

The results of the full kinetic analysis revealed that the autocatalytic reaction is in fact concentration dependent. Specifically, the replication facilitated by the ternary complex [**2**·**3**·**1**] in pathway III contributes 46% at the most, and only if preformed template was added at $t = 0$. Formation of **1** was found to proceed primarily through [**2**·**3**] pathway II. Following these results, the Rebek laboratory developed[113] the system further and increased its efficiency by changing the naphthyl spacer to a bisphenyl linker.

The extension decreased successfully the contribution from the binary reactive complex pathway, and the replicator exhibited a sigmoidal reaction profile.

In 1997, Wang and Sutherland reported[114] the design and experimental demonstration of a self-replicating system based on the Diels-Alder reaction between a maleimide **4** as the ene and cyclohexadiene **5** as the diene (**Figure 1.21**). The reaction of these components in CD_2Cl_2 to form template **6** exhibited a sigmoidal reaction profile. The self-replicating nature of their system was confirmed by examining the same reaction in the presence of preformed template **6**. The authors confirmed the reliance of the system on molecular recognition by demonstrating that when either of the reaction partners is exchanged for an analogue lacking the recognition site, the reaction proceeds at a much slower rate. While the authors undertook analysis of the kinetic and thermodynamic processes governing the system, no discussion of the stereochemical features of the self-replicator was provided. Specifically, both diene **5** and template **6** are chiral, and the reaction of **5** with **4** can result in four different diastereoisomers (two *endo* and two *exo* products). Instead, the authors assigned the observed product as *endo*-**6**, despite providing no analytical evidence to support this assignment.

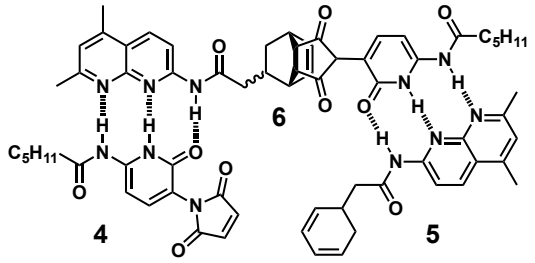


Figure 1.21 A Diels-Alder reaction based replicator reported by Wang and Sutherland. Reaction between ene **4** and diene **5** produces template **6**, capable of templating its own formation *via* the ternary catalytic complex [**4**·**5**·**6**], mediated by hydrogen-bonding recognition. Scheme adapted from Ref. 114.

The possibility of homochiral and heterochiral self-replication presented by the Wang and Sutherland replicator inspired Kiedrowski and co-workers to undertake[115] a significantly more detailed mechanistic and stereochemical study on a variation of this replicating system. The authors replaced the heterocyclic recognition sites on the original reaction components with an amidopyridine and a carboxylic acid (**Figure 1.22**), first reported[116] by Hamilton, to give a new ene **7** and a diene **8**. Diene **8** was synthesised in the racemic form (*rac*-**8**) as well as the separate enantiomers (*R*-**8** and *S*-**8**). Initially, the authors examined the reaction between *rac*-**8** and **7a** and also the methyl-substituted variant **7b**. The reaction profiles for the formation of *rac*-**9a** and *rac*-**9b** products exhibited a lag period, which shortened dramatically in the presence of the corresponding preformed racemic template. Through kinetic fitting of the experimental

NMR data, the authors were able to establish that their replicator design retained its replication efficiency (p = 0.8), similar to that reported by Wang and Sutherland.

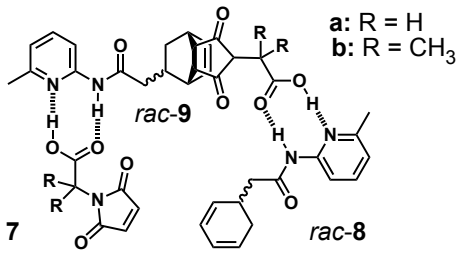


Figure 1.22 A Diels-Alder replicating system inspired by the Wang and Sutherland replicator reported by von Kiedrowski and co-workers. The reaction components, ene **7** and diene **8** assemble with the template **9** in a ternary catalytically-active complex [**7**·**8**·**9**] *via* hydrogen-bonding mediated recognition. Scheme adapted from Ref. 115.

In the next step, the authors undertook a comprehensive kinetic analysis of the homo- and the heterochiral reaction pathways. Specifically, the reaction of each diene enantiomer, *R*-**8** and *S*-**8**, with the ene **7** was examined in the absence of template, followed by analysis in the presence of the enantiopure template *R*-**9a**. The kinetic results showed that the template exerted a similar catalytic effect on both pathways, confirming that both homo- and heterochiral catalytic pathways are effective in the system.

In order to develop a better understanding of the behaviour driving the replicating system, Kiedrowski and co-workers performed a computational analysis. This analysis revealed two conformational families, which differ in the position of the carboxilic acid group relative to the C=C bond (both *endo*). The authors analysed the energy profiles for the various structures present during the reaction of **7**/**7a** (for the two conformational families). The calculations showed that the most significant difference between the homo- and heterochiral pathways is at the transition state. In particular, autocatalytic transition states were determined to have a lower energy than the crosscatalytic pathways leading to the heterochiral products. Moreover, the product duplexes were always found to be more strained and thus less stable than the corresponding ternary complexes. Therefore, the authors rationalise that the near exponential growth observed in this replicating system can be explained in terms of the conformational constraints in the product duplexes.

Diels-Alder reaction has been exploited[117–122] as the ligation step in recognition-mediated reactions and replicating systems also in the Philp laboratory. In their work, Philp and co-workers have utilised the Diels-Alder reaction in the design of two structurally-similar families of replicators as platforms for investigating the effect of structural variation on the efficiency of replication and other recognition-mediated channels in each system. Specifically, the first series of replicators (**Figure 1.23**) in-

vestigated[117,118,121] a system based on a reaction between a furan and a maleimide. The furan was equipped with an amidopyridine recognition site at 2- or 3- position, complementary to a carboxylic acid recognition site on the maleimide. Additionally, the authors also explored the effect of varying the number of methylene spacers between the recognition and the reactive site in the furan (m = 1 or 2) and the maleimide (n = 1 or 2) components. Reaction of the furan and maleimide produces two diastereoisomeric templates, *endo* and *exo*. Overall, the work examined eight different reactions (labeled I to IV for each 2- and 3-substituted furan) (**Figure 1.23**).

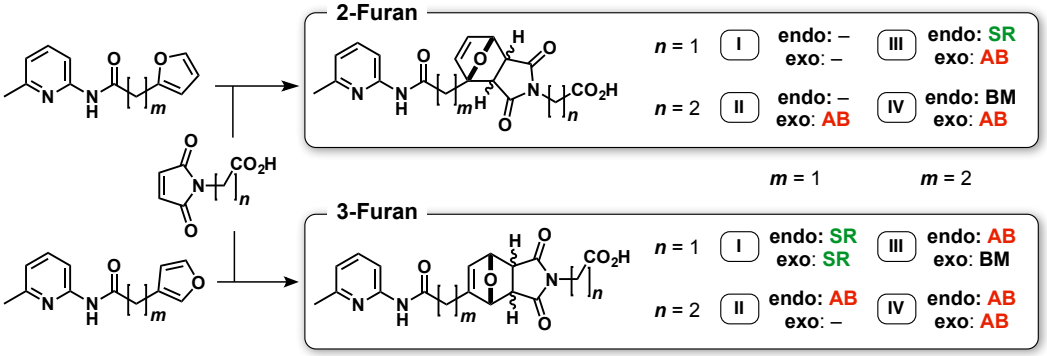


Figure 1.23 Series of structural modifications explored by Philp and co-workers. Reaction of 2- and 3-furan derivatives equipped with an amidopyridine recognition unit with a maleimide is mediated by the presence of a carboxylic acid, affording a set of potential replicators (I to IV in 2- and 3-furan susbstituted system) (35 °C, $CDCl_3$). Diels-Alder reaction of the furan and maleimide produces *endo* and *exo* diastereoisomeric templates. SR denotes self-replicating reactivity (highlighted in green), AB represents reaction through binary reactive channel (red) and BM represents no recognition-mediated reactivity. –Denotes no significant conversion. Figure adapted from Ref. 120.

Comprehensive kinetic analyses of the various reactions showed that out of the 16 possible templates, only two *endo* templates and a single *exo* template are capable of templating their own synthesis (**Figure 1.23**, green SR). Additional seven templates were found to form *via* a reactive binary complex pathway (**Figure 1.23**, AB). The remaining templates showed no measurable recognition-mediated reactivity. Given the data collected, the authors reasoned that a system with a high degree of conformational freedom is more likely to react through the binary complex pathway preferentially. Therefore, developing a highly efficient self-replicating system necessitates a certain degree of rigidity as well as a suitably 'open' arrangement of the recognition sites in space—a requirement for the formation of catalytically-active ternary complexes. Interestingly, the maleimide (n = 1) employed in this study is identical to the maleimide **7** studied by von Kiedrowski, where cyclohexadiene is employed in place of the furan component. The structurally most similar system in this work to that examined by von Kiedrowski are the two systems labelled as I. Remarkably, the 2-furan system I showed

no reactivity whereas the 3-substituted furan system I had the capacity to self-replicate, albeit less efficiently than the Kiedrowski replicator, which exhibited exponential growth. Comparison of these structurally-similar systems highlights that even very small structural changes can exert a remarkably drastic effect on the ability of a system to take part in recognition-mediated reactions, thus, showing the challenging nature of designing self-replicating systems—a process that can benefit from computational predictions.

In later work, Philp and co-workers examined[119,120] a second, complementary series of reactants, with an amidopyridine unit now located on the maleimide and the carboxylic acid on the furan component. Again, reaction of four furans and two maleimides affords 16 diastereoisomeric reaction products (**Figure 1.24**). Through kinetic analyses, the authors found that the ability to self-replicate was limited to 6 templates, all based on the 3-furan motif (**Figure 1.24**, green), whereas templates based on 2-substituted furan formed through a binary complex channel or showed no recognition-mediated reactivity. Again, spacer length as well as the position of the linker bearing the recognition site on the reaction component were shown to effect the ability to replicate. The thorough analyses of these two families of structurally-related replicators highlights the limit on the structural window that permits self-replication—a window that can be probed by computational methods (*e.g.* transition state modelling).

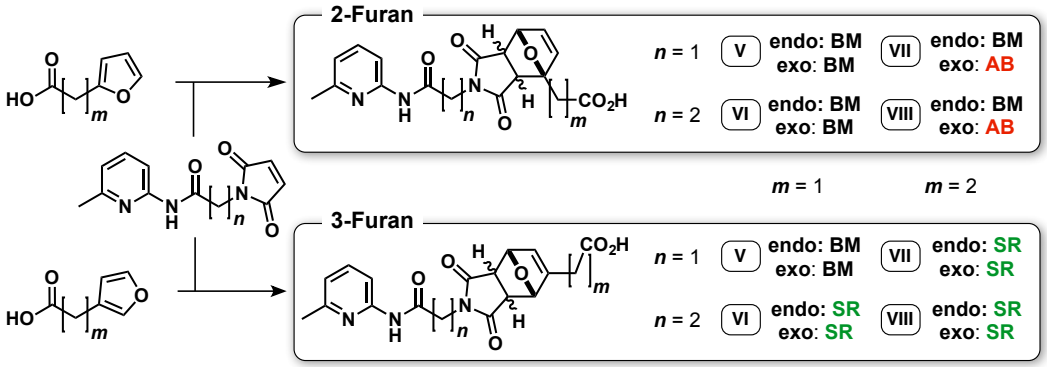


Figure 1.24 Series of structural modifications explored by Philp and co-workers. Reaction of 2- and 3-furan derivatives equipped with an carboxylic acid with a maleimide is mediated by recognition of the amidopyridine group on this element, affording a set of potential replicators (V to VIII in 2- and 3-furan substituted system) (35 °C, $CDCl_3$). Diels-Alder reaction of the furan and maleimide produces *endo* and *exo* diastereoisomeric templates. SR denotes self-replicating reactivity (highlighted in green), AB represents reaction through binary reactive channel (red) and BM indicates no recognition-mediated reactivity. Figure adapted from Ref. 120.

The Philp laboratory has utilised[123] the simple maleimide **7a** (n = 1) equipped with a carboxylic acid group in the design of a replicator exploiting 1,3-dipolar cycloaddition reaction between this maleimide as the dipolarophile and an azide **10** as the 1,3-dipole

(**Figure 1.25a**). This design was inspired[124] by previous work, which explored the same reaction but utilised a maleimide with an additional —CH_2 spacer. The structural flexibility of the longer maleimide resulted in the preference of the system for the binary reactive complex pathway and the authors envisaged that reduction in the spacer length would ensure sufficient structural rigidity, necessary to promote self-replication. Kinetic studies showed that in the altered design, template **11** was capable of self-replication mediated by the ternary complex [**7a**·**10**·**11**]. The resulting template duplex [**11**·**11**], however, proved to be very stable, limiting the autocatalytic turnover in the system as a result of significant product inhibition ($p = 0.4$).

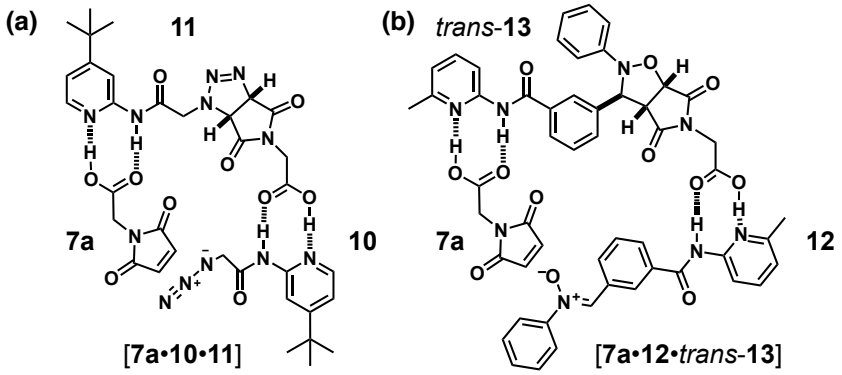


Figure 1.25 **(a)** A replicating system exploiting 1,3-dipolar cycloaddition reaction between a maleimide **7a** and an azide bearing component **10**. The product **11** formed by the reaction of these components is capable of templating its own synthesis *via* the formation of the ternary complex [**11**·**11**]. **(b)** Replicator design exploiting the reaction between maleimide **7a** and nitrone **12**. The reaction of these components produces two diastereoisomeric products *trans* and *cis* but the *trans* template **13** only is capable of templating its own formation, that is self-replicating *via* [**7a**·**12**·**13**] complex. Figure **(a)** adapted from Ref. 123 and **(b)** from Ref. 125.

Philp and co-workers have further developed[125] the 1,3-dipolar cycloaddition reaction-based replicating systems by employing a nitrone as the 1,3-dipole. Reaction of nitrone-bearing molecules with maleimides results in two diastereoisomeric products, *trans* and *cis*[a], with noticeably different geometries. Replicating systems employing such reactions thus provide a platform for investigating the transfer of stereochemical information. The authors investigated the 1,3-dipolar cycloaddition reaction of maleimide **7a** with nitrone **12** (**Figure 1.25b**). Molecular recognition in the system was provided again by the association of a carboxylic acid moiety with the 6-methylamidopyriidne group described previously. In this particular system, the *trans* diastereoisomer of the template **13** only possesses the open conformation necessary for

[a]The *cis* and *trans* notation is used[126] to reflect the relative configuration of the three protons located on the bicyclic ring structure formed in the cycloaddition reaction. In the *trans* cycloadduct, the proton derived from the nitrone is located on the opposite of the two protons originating from the maleimide component. In the *cis* cycloadduct, the protons derived from the nitrone and maleimide components are located on the same face of the fused ring system.

successful docking of the building blocks in the ternary catalytic complex [**7a**·**12**·**13**], driving the self-replication cycle. Initially, the authors examined the reaction of nitrone **12** with a recognition-less methyl ester of maleimide **7a** ($CDCl_3$, 10 °C). In the absence of recognition, this reaction afforded the recognition-disabled *trans*-**13** and *cis*-**13** analogues in a ratio of 3. The same reaction with maleimide **7a** performed much more efficiently (p = 0.9), exhibiting a sigmoidal reaction profile and producing the [*trans*]/[*cis*] diastereoisomeric products in a ratio = 6. Undertaking the reaction in the presence of preformed *trans*-**13**, but not *cis*-**13**, resulted in the disappearance of the lag period and an increase in the diastereoselectivity for the *trans* product ([*trans*]/[*cis*] = 9). Through these kinetic experiments, the authors confirmed that the self-replicating template is capable of transmitting structural information successfully *via* the ternary complex pathway.

While the nitrone-based self-replicating design performed efficiently, only a modest amplification in the reaction rate and the diastereoisomeric ratio were achieved. Exploiting the information available from the previous structure *versus* reactivity studies and computational modelling, Philp and Kassianidis designed[127] a structurally-optimised replicator, exploiting structural modifications in the design of both the nitrone and the maleimide element, intended to disfavour the reactivity *via* the binary complex pathway. This new replicator (**Figure 1.26**) was formed by the reaction of a more extended phenylacetic acid maleimide **M1** with nitrone **14**. These structural changes facilitated the formation of a significantly more stereoselective replicator *trans*-**15**, capable of achieving a [*trans*]/[*cis*] ratio of 115 even in the absence of preformed template. The reaction was further biased towards the formation of the *trans* diastereoisomer by addition of preformed template of *trans*-**15** and by reduction of the reaction concentration.

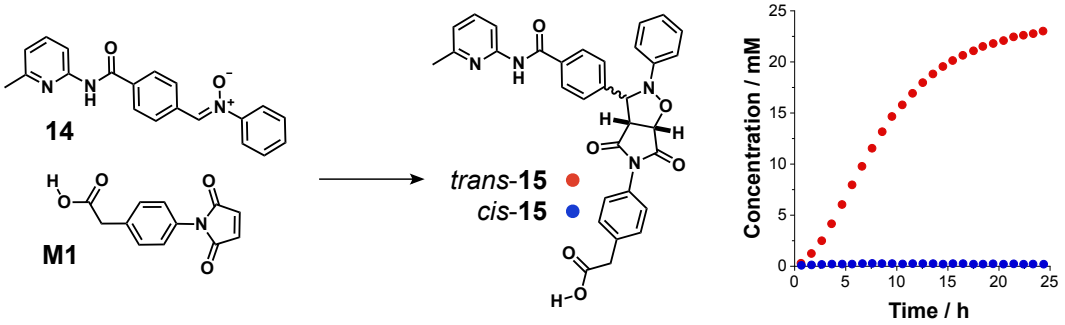


Figure 1.26 Reaction between nitrone **14** and maleimide **M1**, and the corresponding concentration *vs* time profile ($CDCl_3$, −10 °C) produces a highly diastereoselective self-replicator *trans*-**15** (red circles) capable of amplifying itself at the expense of the *cis* cycloadduct (blue circles) (ratio of [*trans*]/[*cis*] > 125:1).

The presented examples of self-replicating systems illustrate the marked progress that has been achieved in the design and implementation of minimal self-replicating

systems based on small synthetic molecules. Some of these examples have been extended by Rebek[128,129] and Philp[127,130,131] into interconnected networks incorporating formation of more than a single replicator, as well as replicators functioning in the reciprocal sense. Philp and co-workers have combined[130] the 3-furan system described in **Figure 1.24** (m = 2, n = 1) with the efficient self-replicating system shown in **Figure 1.26** to construct a network comprised of four building blocks: **14**, **M1**, **16** and **17** (**Figure 1.27**). Reaction of these components produces three established independent self-replicators *trans*-**15**, *endo*-**18** and *exo*-**18**. Alternatively, combination of the reaction components can produce also two mutually complementary templates, *trans*-**19** and *exo*-**20**, capable[127] of taking part in reciprocal replication. Examination of the reaction of all four components ($CDCl_3$, 25 °C) revealed that the system makes *trans*-**15** preferentially (accounting for over 70% of the product pool).

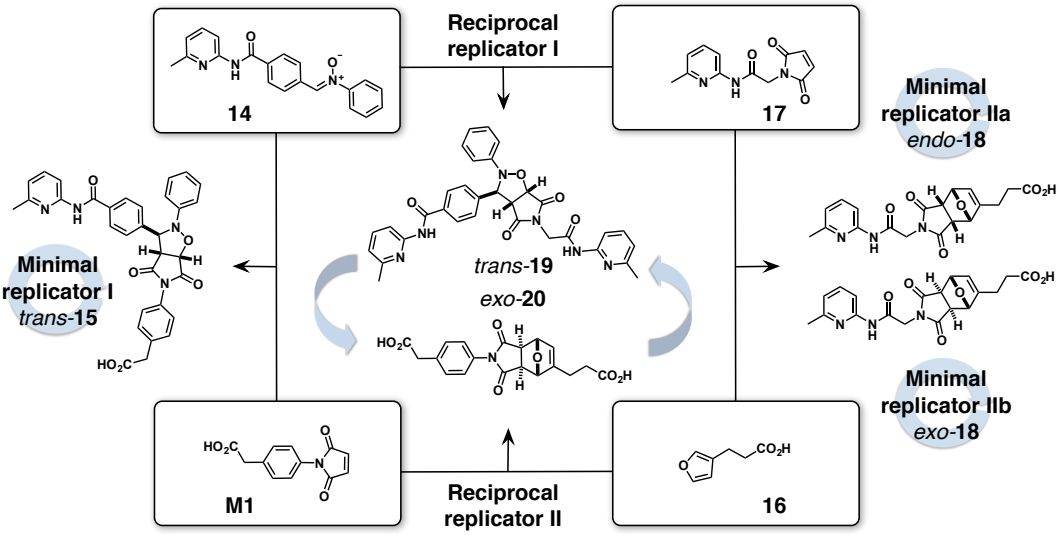


Figure 1.27 A multicyclic replication network built from two maleimides, **M1** and **17**, nitrone **14** and a furan **16**. Reaction of these components produces three minimal self-replicating systems *trans*-**15**, *endo*-**18** and *exo*-**18** and two reciprocal products, *trans*-**19** and *exo*-**20**. Strong preference of the system for the formation of *trans*-**15** imposed a limit on the instructability of the system with preformed templates. Cycloadducts exhibiting no recognition-mediated activity are omitted for clarity. Figure adapted from Ref. 130.

Initially, the authors envisaged that the network could be directed towards increased production of a specific template by addition of preformed templates. However, the strong bias of the system for *trans*-**15** as a result of the uneven replication efficiencies resulted in a network that is fairly insensitive to addition of any template. Using kinetic simulations, the authors investigated a number of conditions, varying the catalytic efficiencies for the auto- and crosscatalytic replicators, as well as the amount of preformed template used as input. Interestingly, the authors showed that the components replicating in the reciprocal sense are more responsive to instruction, compared to the autocatalytic

templates, most likely as a result of the mutually-reinforcing nature of these reciprocal templates.

Aided by computational modelling and kinetic simulations, the research directed at investigating self-replicating systems has progressed dramatically, over the last 30 years, from the first examples of minimal, often inefficient, replicating systems to significantly more varied and interconnected networks exploring the interplay between various recognition and reaction features. The numerous experimental implementations of self- and reciprocal replicating systems now available in the literature complement the theoretical models, leading to a better understanding of the principles governing the reactivity and information transfer in these systems. The interconnectedness of the components in these systems endows them with the capacity to respond to stimuli, allowing complex function to emerge. In particular, the resulting chemical systems express the complex phenomena of self-replication, and often also other functionality such as Boolean logic operations, error-correction and stereospecific replication. Understanding the system-level behaviour in interconnected networks, and, thus, also the possibility of harnessing these complex replicating systems in the construction, selection and amplification of higher order assemblies, necessitates the understanding of the kinetic and thermodynamic forces driving the recognition and reaction processes behind the individual components in a system.

1.5 Dynamic covalent chemistry

Nature presents scientists with an abundance of complex systems. However, the study of these natural complex systems is often limited by the sheer number of components comprising them, the high level of their interconnectedness, and the insufficient tools currently available for their analysis. Yet, the desire to understand and study complex systems remains. Two main approaches can be adopted in order to circumvent the overwhelming complexity of natural systems. Firstly, the top-down approach, directed at separating a particular complex system into smaller, more manageable subsystems (the process can be repeated to create further subsystem levels), which can be examined in isolation. Alternatively, a bottom-up approach to building complex systems can be employed by combining different individual parts to form a larger system. Chemists in particular, often seek to design complex chemical networks using the bottom-up approach, with specific features in mind, in order to enable them to test a specific hypothesis.

Over the years, one method has proven to provide efficient and direct (bottom-up) protocol for constructing synthetic interconnected networks—dynamic covalent chem-

istry (DCC).[132–135] The term dynamic covalent chemistry should be distinguished[135] from the term dynamic combinatorial chemistry,[136,137] of which it is a subset, together with dynamic non-covalent chemistry. The concept of dynamic combinatorial chemistry and the associated general requirements were first identified[138] by Huc and Lehn in 1997. The authors made the distinction between dynamic combinatorial chemistry as a conceptually different approach to that of combinatorial chemistry itself. The dynamic combinatorial approach is driven by reversible reactions, covalent or non-covalent, between a set of components, spontaneously forming a virtual library of all the possible combinations of products (as permitted by the structural and recognition features on the components themselves). Such a virtual library, if exposed to a target, should result in the selection (*i.e*. amplification) of a specific component with the features allowing the most optimal interaction with the target. Virtual combinatorial libraries were identified as a potential tool for the discovery of a substrate for a particular receptor (casting) or the construction of a receptor for a particular substrate (molding) (**Figure 1.28**).

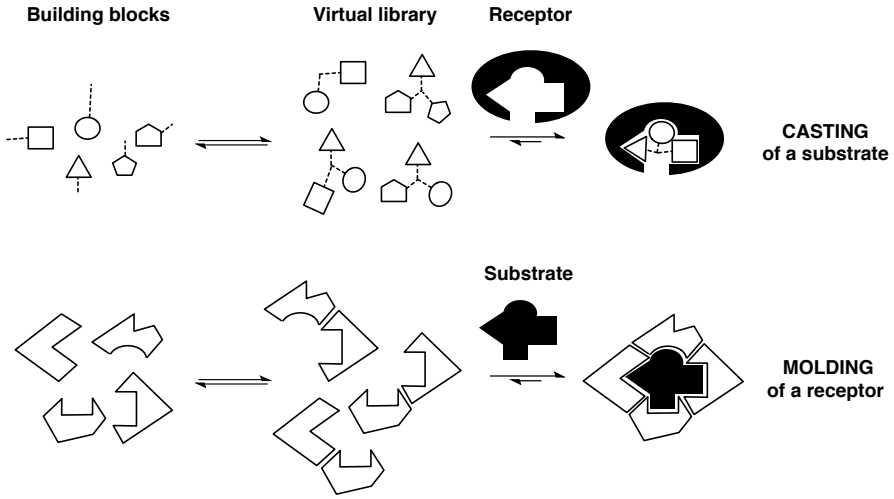


Figure 1.28 Formation of virtual combinatorial libraries from a set of building blocks as proposed by Huc and Lehn in 1997. A diverse array of virtual library components (centre) is formed either *via* covalent or non-covalent linkage. In the presence of a binding partner (a receptor or a substrate), the virtual library re-equilibrates so as to select components capable of the most optimal interaction with the binding partner. Schematic representations of the receptor-induced casting process and substrate driven molding process from a virtual library illustrate the selection method. Figure adapted from Ref. 138.

The DCC method exploits the reversible covalent bond formation between various building blocks equipped with compatible reactive sites. The benefit afforded by the robustness of the reversible bond formation coupled with the general combinatorial approach facilitates formation of structurally-diverse dynamic covalent libraries (DCLs) of dynamically-exchanging components under thermodynamic control. The equilibrium distribution of such libraries can be examined to reveal the most thermodynamically-stable product distribution, in the absence of any binding partners. As a result of the

reversibility of bond formation in the DCC approach, the constructed dynamic systems exhibit a capacity for error-checking and error-correction. The products formed through the reversible covalent bonds, while potentially slower to form, are often more stable than supramolecular assemblies relying on non-covalent interactions. For covalent bond forming reactions to be suitable for use in the DCC method, the employed reaction products should be stable enough to allow detection of the library components but also be formed fast enough to permit equilibration on a reasonable time-scale. Ideally, the bond forming reactions should proceed under mild reaction conditions, and, therefore, be compatible with a wide range of functional groups. Generally, the type of reaction employed to generate a dynamic covalent library involves either exchange of one reaction partner with another or formation of new dynamic bonds in a product that combines the building blocks. A selection of reactions commonly utilised in DCC is illustrated in **Figure 1.29**. Ultimately, with these features, the DCC approach can facilitate the formation of large supramolecular structures, identification of receptors, sensors and bioactive compounds that might be much less accessible using the traditional synthetic techniques.

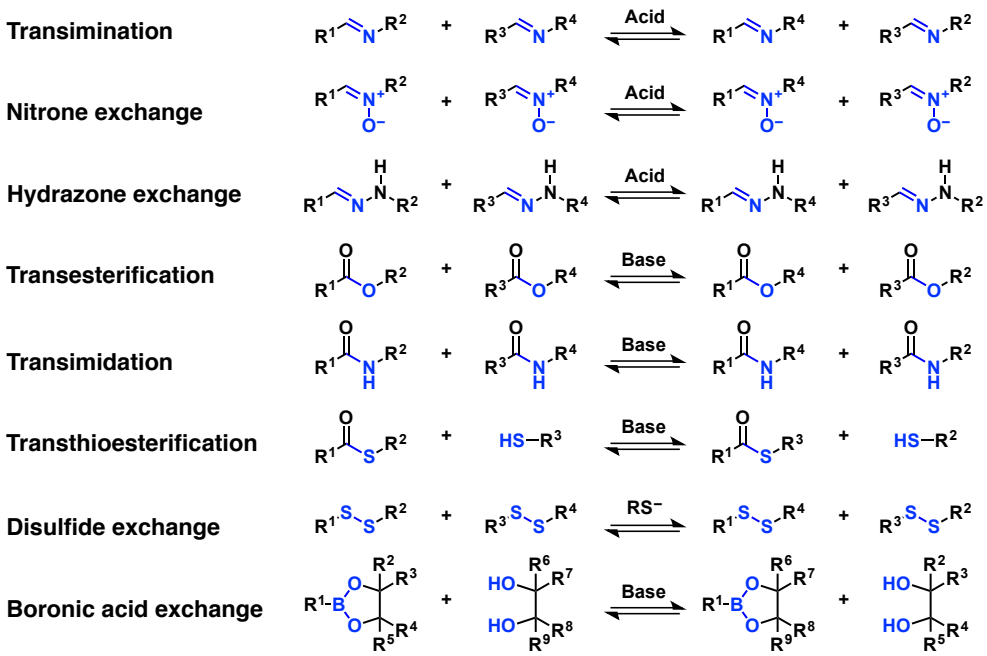


Figure 1.29 Example reversible covalent reactions employed in dynamic covalent chemistry.

As introduced[138] by Huc and Lehn, dynamic covalent libraries can be instructed by the addition of an external stimulus, capable of altering the thermodynamic distribution of a library. The added stimulus, for example a template, with the capacity to recognise and thus interact with the species present in the library through non-covalent recognition processes, results in library redistribution so as to reflect the new most

thermodynamically-stable state of the entire system. Ideally, the compounds present at higher concentration are those capable of engaging in stronger recognition. The instructed library will amplify preferentially, relative to the uninstructed library, formation of components with the highest affinity for the added template. In this manner, DCC offers an efficient approach to screening of a large number of virtual components (all potential combinations that can be synthesised *in situ* from the library building blocks) and discovering those components that bind the most strongly. In practice, the strength of binding can be determined by comparing the concentrations of library members in the absence of instruction, relative to the concentrations observed in the presence of the template.

1.5.1 Dynamic covalent libraries under thermodynamic control

In addition to establishing the notion of dynamic combinatorial chemistry, Huc and Lehn also presented[138] an experimental implementation of this concept by examining the formation and selection of carbonic anhydrase (CA) II inhibitors formed from a virtual library established from amines and *para*-substituted sulfonamide aldehydes. Using HPLC as a detection method, the authors exploited $NaBH_3CN$-mediated reduction of the imines to the corresponding amines in order to freeze the equilibrated state of the library. In the next step, the same procedure was repeated for the examination of the library distribution after receptor-induced re-equilibration. While the presence of CA exerted no effect on some of the library members, the authors were able to observe some selectivity for compounds structurally-similar to those of known CA inhibitors.

Stoddart and co-workers have exploited[139] the dynamics of imine bond formation for the construction of a mechanically-interlocked Borromean ring **21** (**Figure 1.30**). In this example, where DCC is employed in the context of supramolecular chemistry, the authors used a dialdehyde **22** and a diamine **23** (**Figure 1.30a**). Clearly, in the absence of a template, a number of combinations can be formed by the reaction of these components. However, Stoddart and co-workers found that in the presence of Zn^{2+}, a supramolecular assembly **21** (**Figure 1.30c**) of three mutually-interwoven macrocycles **24** is produced (**Figure 1.30b**). In particular, the assembly of 6 aldehydes **22** and 6 amines **23** is templated by 6 metal ions. The arrangement of these macrocycles in the Borromean ring **21** (**Figure 1.30c**) is such that they can not be separated without breaking of the covalent bonds in one of the macrocycles.

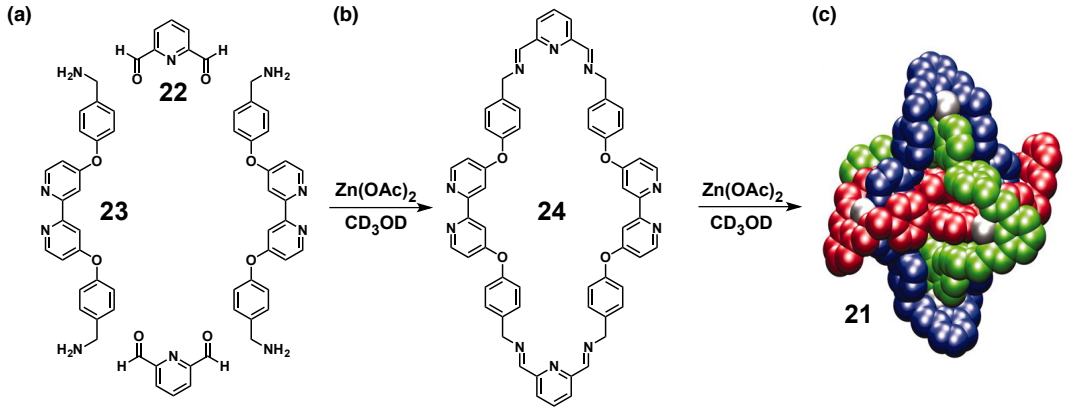


Figure 1.30 **(a)** Aldehyde **22** and amine **23** can react reversibly to form various assemblies, however, in the presence of $Zn(OAc)_2$, these components results in a highly symmetrical interwoven assembly comprised of three **(b)** macrocycles **24**. **(c)** Space-filling representation of the structure of Borromean ring **21** deduced using X-Ray crystallography. The three macrocycles forming the assembly are highlighted in different colours. Metal ions are shown in grey. Figure adapted from Ref. 139.

In another example, of perhaps more traditional dynamic covalent chemistry, Otto and co-workers have successfully identified[140] a catalyst for a Diels-Alder reaction from a dynamic covalent library utilising a transition state analogue (TSA) as the target. The authors have demonstrated[141] previously that disulfide chemistry can be exploited for generation of libraries of macrocycles in water, under mild reaction conditions (pH 7 to 9), in the presence of thiolate. In this work, three dithiol building blocks, **25**, **26** and **27** (**Figure 1.31a**) were employed as the library building blocks. The authors were interested in catalysing the reaction between acridizinium bromide **28** and cyclopentadiene **29** (**Figure 1.31b**), leading to product **30**. Reaction between these components proceeds through a transition state that is very structurally-similar to the product itself, and the authors envisaged that product **30** would serve as a suitable TSA. Mixing of one equivalent of each of the three dithiol building blocks (at pH 8.5, up to five days) with **30** as input, showed a dramatic amplification of two receptors in particular: a homotrimeric racemic receptor **31** and heterotrimeric receptor **32** (**Figure 1.31c**). Interestingly, these two macrocycles were identical to those amplified[142] from a DCL comprised of the same building blocks, instructed with hydrophobic ammonium ions. Prior to investigating the catalytic activities of these two identified receptors, the authors measured the binding affinity of **31** and **32** with the starting material **28** and product **30**, finding that only the homotrimeric receptor **31** binds the product more strongly than the starting material.

Analysis of the Diels-Alder reaction rate between **28** and excess of cyclopentadiene **29**, first in the absence of a receptor and later also in a mixture including 1.2 equivalents of receptor **31** or **32**, showed that, as predicted based on the binding studies, only the

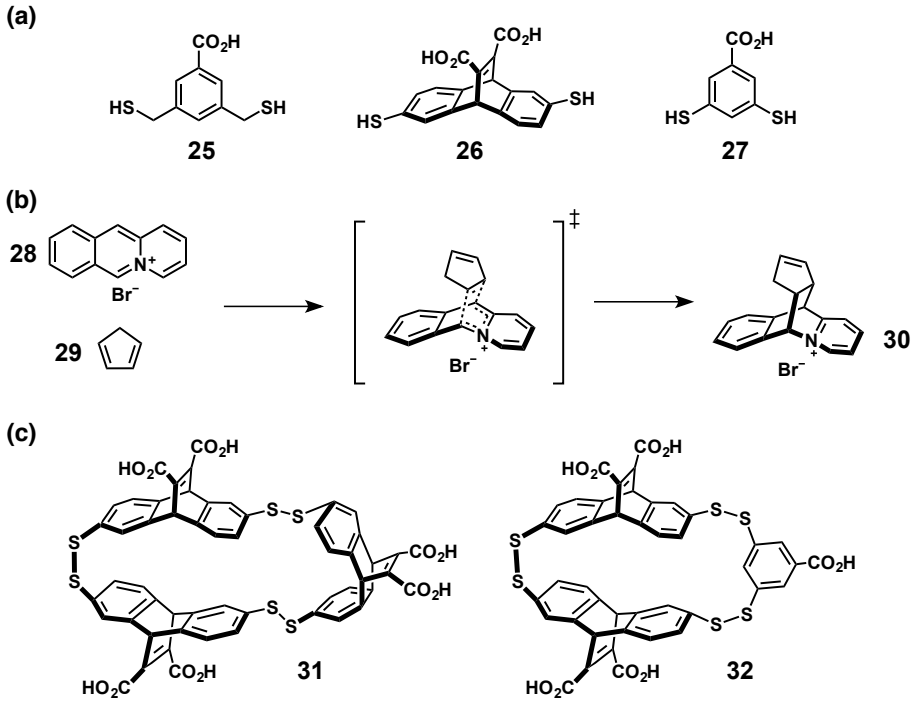


Figure 1.31 **(a)** Three dithiol building blocks, **25**, **26** and **27**, capable of forming a dynamic covalent library of macrocycles *via* disulfide exchange in water. **(b)** Diels-Alder reaction between acridizinium bromide **28** and cyclopentadiene **29** proceeds through a transition state that is structurally-similar to the reaction product **30**. **(c)** A DCL comprised of dithiol building blocks **25**, **26** and **27** instructed with **30** as a transition state analogue results in the amplification and selection of a homotrimeric racemic receptor **31** and heterotrimeric racemic receptor **32**. Figure adapted from Ref. 140.

presence of **31** resulted in rate acceleration, albeit a modest one. The low catalytic activity was rationalised by the limited turnover as a result of receptor **32** inhibition by the resulting product. Ultimately, despite the modest catalytic efficiency, the study illustrates the potential utility of DCC approach to catalyst discovery.

The Philp laboratory has become interested in utilising recognition-driven reaction processes as a tool for selection and amplification of library components in dynamic covalent libraries. In order to address the need for reversible reactions compatible with non-polar organic solvents, suitable for the recognition-mediated reactions employed within the group previously and at the same time presenting the possibility of further chemical transformations, Philp *et al*. explored[143] nitrone exchange. Examining an equimolar (10 mM) solution of two nitrones (**Figure 1.32**) **33** and **34**, each equipped with an amidopyridine recognition unit, in the presence of a catalytic amount of 4-fluorophenylhydroxylamine **35**, the authors were able to establish that the nitrone exchange has reached equilibrium after 48 hours. The equilibrium mixture contained the four nitrones in a 1 : 1 : 1 : 1 ratio and the same equilibrium was reached when the experiment was repeated starting from **36** and **37**. In this mixture of four nitrones,

nitrone **36** bears no recognition sites, while nitrone **37** bears two amidopyridine recognition elements, and, thus, is the only component in the library capable of binding a dicarboxylic acid. When the exchange of nitrones **33** and **34** was examined in the presence of a dicarboxylic acid template **38**, a clear selectivity for nitrones **36** and **37** ([**36** + **37**]/[**33** + **34**] = 3.9) over the two other nitrones was observed. The degree of amplification was found to be dependent on the amount of diacid template added. While the authors were able to establish through density functional theory (DFT) calculations that the amplification of nitrone **37** is limited by a mismatch between the length of the distance between recognition sites on the diacid template **38** and nitrone **37**, the study established unambiguously for the first time that diaryl nitrone can undergo exchange in non-polar solvents—establishing nitrone exchange as a viable tool for DCC.

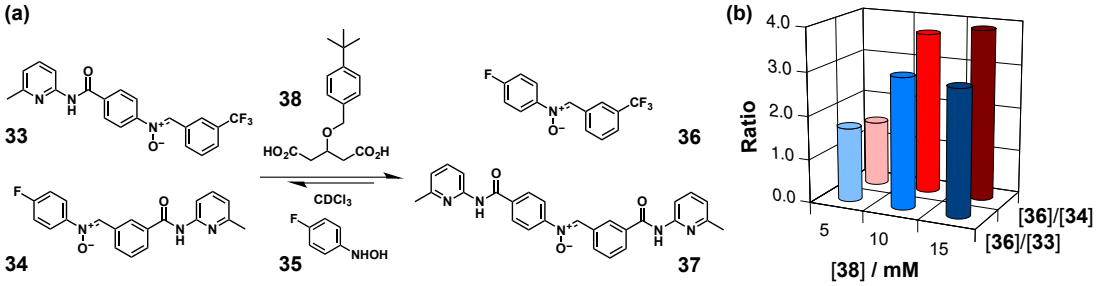


Figure 1.32 **(a)** Dynamic covalent library examining nitrone exchange in a non-polar solvent. Under equilibrium conditions, the four possible exchange products, **33** to **37**, are present at nearly identical concentrations. Stabilisation of **37** by binding to a diacid template **38** resulted in perturbation of the equilibrium and amplification of nitrones **36** and **37**. **(b)** Ratio of nitrones [**36**]/[**33**] (blue bars) and [**36**]/[**34**] (red bars) as a function of concentration of diacid template **38**. Scheme adapted from Ref. 143.

An interesting example of a dynamic library was presented[144] by Buryak and Severin, where instead of utilising the library for the identification of a specific sensor, the entire library of dye-metal complexes was employed for sensing. The authors exploited the fact that a particular library composition depends on its environment, opening up the possibility of transducing library distribution into a signal output. In this work, the authors specifically focused on the identification of dipeptides in water. Utilising three commercially available dyes, **39**, **40** and **41** (**Figure 1.33**) and copper and zinc salts, the authors first established that upon mixing, these components are in a dynamic equilibrium with a small proportion of the uncomplexed species. The capacity of this library for dipeptide sensing exploited the fact that dipeptides can form stable complexes with Cu^{2+} and Ni^{2+}, thereby displacing these ions from the dyes, a process which in turns leads to library re-equilibration and an increase in the concentration of the free dyes in solution. The library exhibited efficient sensing even with structurally-similar dipeptides—while the UV-Vis spectra (**Figure 1.33**) were fairly similar to each

other, a chemometric technique[b] enabled successful differentiation even between closely related anylates.

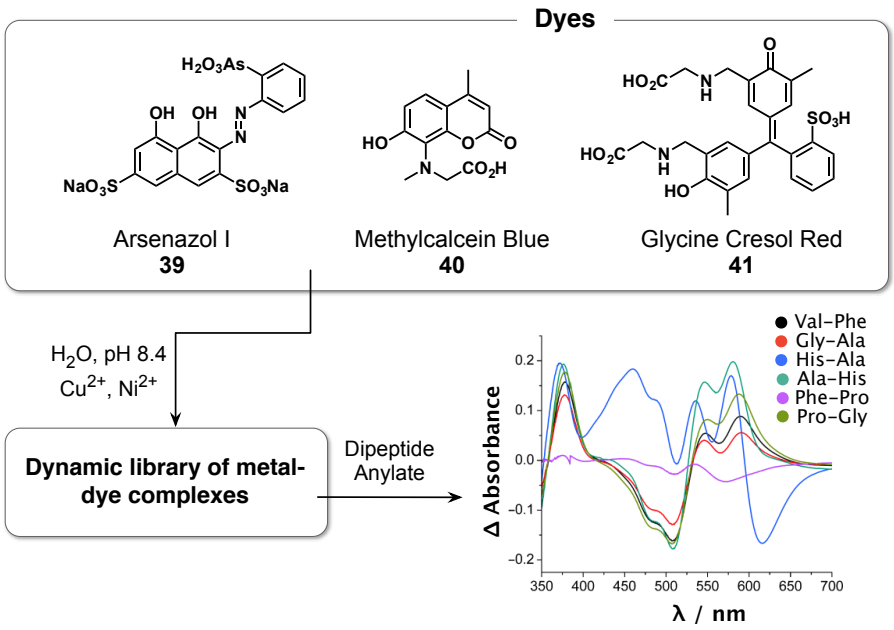


Figure 1.33 A dynamic library of metal-dye complexes is assembled using three dyes, **39**, **40** and **41** with copper and zinc metal ions in aqueous buffer. In the presence of various dipeptides as anylates, the library acts as a sensor for their detection by UV-Vis spectroscopy. Figure adapted from Ref 144.

Ashkenasy and co-workers have extended[145] the light-induced replication protocol described previously to a dynamic peptide system replicating reversibly under thermodynamic control (**Figure 1.34**). This system incorporates reversible formation of two peptide products, $\mathbf{R^1}$ and $\mathbf{R^2}$, from electrophilic fragments $\mathbf{E^1}$ (**Figure 1.34**, green), $\mathbf{E^2}$ (**Figure 1.34**, purple) and a shared nucleophilic peptide building block **N** (**Figure 1.34**, grey). Replication in the system was made reversible by substitution of the reactive cysteine residue employed previously, with a thioglycolic acid at the N-terminus of the nucleophilic fragment **N**. Ligation reaction between the smaller peptide fragments produced a thioester bond at a central position in each peptide template, allowing for reversible *trans*-thioesterification. Templates $\mathbf{R^1}$ and $\mathbf{R^2}$ differ only in the electrophilic fragment, where $\mathbf{R^1}$ contains a glutamate residue and $\mathbf{R^2}$ incorporates a lysine residue at position e (13) within the heptad repeat, directly opposite a lysine residue at position g'. As a consequence of the repulsive electrostatic interactions between the lysine residues, $\mathbf{R^2}$ is not capable of forming the stable catalytically-active intermediates, required for efficient self-replication. In the absence of external triggering, $\mathbf{R^1}$ only is capable of efficient replication through the dimeric template-mediated thioesterification (**Figure 1.34**).

[b]Chemometric approach utilises data analysis for obtaining information from chemical systems.

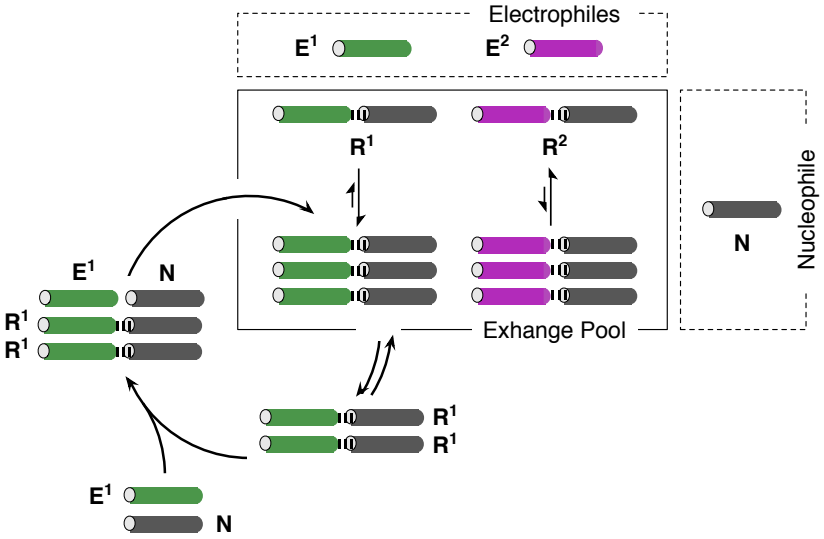


Figure 1.34 Replicating peptide network under thermodynamic control designed by Ashkenasy and co-workers. Two electrophilic components, $\mathbf{E^1}$ (green) and $\mathbf{E^2}$ (purple) can react with a nucleophilic component **N** (grey) to form reversibly peptide templates $\mathbf{R^1}$ and $\mathbf{R^2}$. $\mathbf{R^1}$ only is capable of forming catalytically-active template assemblies, allowing it to establish efficient self-replicating cycle. Adapted from Ref. 145.

In the absence of template, reaction of $\mathbf{E^1}$ with **N** exhibits a sigmoidal reaction profile. As expected for a self-replicating system, the rate of template formation was found to correlate with the quantity of preformed template added to the reaction mixture. The reaction order p (0.65) determined for replicator $\mathbf{R^1}$ is similar to the values of reaction order observed in peptide systems replicating non-reversibly. Interestingly, doping with preformed template **T**, containing the native non-reversible peptide linkage, revealed that this template crosscatalyses the formation of $\mathbf{R^1}$. Behaviour of the small DCL (**Figure 1.34**) comprised of the three building blocks was examined in the absence and in the presence of instructing chemical trigger, *e.g.* template $\mathbf{R^1}$ or **T**. The network expressed interesting behaviour, where concentration of $\mathbf{R^1}$ peptide continued to increase throughout the reaction, as expected based on its replication efficiency, while an initial increase in the concentration of $\mathbf{R^2}$ product was followed by its decomposition to the starting materials, reaching a dynamic equilibrium. Decomposition of $\mathbf{R^2}$ increased the transient concentration of the shared building block **N**, thereby allowing $\mathbf{R^1}$ to be formed at the expense of $\mathbf{R^2}$. Instructing the DCL with $\mathbf{R^2}$ template has resulted in delayed equilibration towards the formation of $\mathbf{R^1}$. Utilising an external, preformed template ($\mathbf{T^{2C}}$), incorporating two glutamate residues, as input, the authors demonstrated that because of the these charged residues complementary to the lysine residues in $\mathbf{R^2}$, this template forms stabilising heteromeric complexes with $\mathbf{R^2}$. Therefore, competition between $\mathbf{R^1}$ and $\mathbf{R^2}$ in the presence of $\mathbf{T^{2C}}$ alters the product pool distribution and near even concentration of these two products is observed at equilibrium.

The possibility of altering the behaviour of the network through the application of light as a stimulus was examined in the presence of added photocleavable template $\mathbf{T^{1Nv}}$. Application of light to the reaction mixture resulted in the cleavage of the protecting photocleavable moiety, exposing the lysine residue (e13) on this template, enabling it to crosscatalyse the formation of $\mathbf{R^1}$. In this work, Ashkenasy and co-workers demonstrated experimentally for the first time that reversible peptide replication is possible and that the product distribution within an interconnected peptide network under thermodynamic control depends on both the catalytic efficiency of each template species as well as their thermodynamic stability. These results are of clear relevance for the understanding of the process of molecular evolution and formation of metabolic networks, and might possibly be extended to networks exhibiting chemical evolvability in the future.

As one of the first examples examining self-replication in a reversibly formed systems, Kiedrowski and Terfort explored[146] amidinium-carboxylate salt bridge as an alternative to nucleotide base pairing recognition in order to drive self-replication of a small molecule-based synthetic system. The authors investigated the condensation reaction of several amines and aldehydes (**Figure 1.35**). The formed imines were suggested to be suitable templates for the association of the unreacted components in a catalytically-active ternary complex. Analysis of the reaction between amine **42** and aldehyde **43** (DMSO-d_6), to give imine **44** (**Figure 1.35a**), showed that the rate of product formation increases as the amount of preformed template **44** added to the reaction was progressively doubled, following the square-root law.

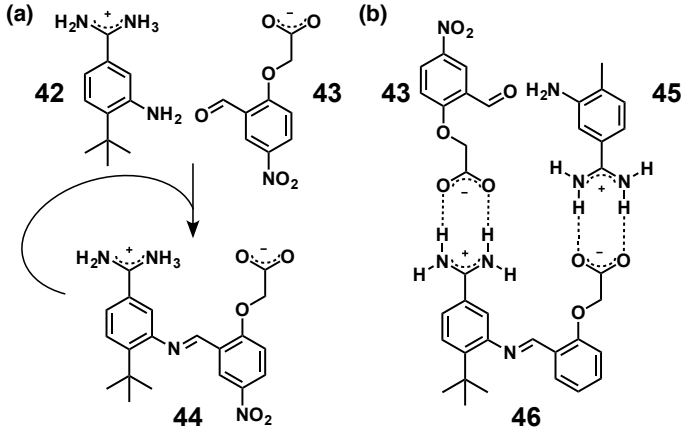


Figure 1.35 Small-molecule synthetic replicating system, exploiting carboxylate-amidinium salt bridge formatiom, investigated by Terfort and Kiedrowski based on condensation of amines and aldehydes. **(a)** Amine **42** and aldehyde **43** form imine **44**, capable of self-replication (limited by product inhibition). **(b)** Crosscatalytic reaction of amine **45** with **43** mediated by imine template **46** showed exponential growth. Counter ions are omitted for clarity. Figure adapted from Ref. 146.

Interestingly, examination of a structurally-similar system where an amine **45** reacted with aldehyde **43** in the presence of imine **46** (**Figure 1.35b**) showed exponential growth,

with $p = 1$. While this system constitutes an example of a crosscatalytic replication, the authors showed that it is possible to overcome product inhibition in small molecule-based replicating systems.

Xu and Giuseppone reported[147] an example of a more complex imine library, where only one component of the exchange pool can be stabilised by formation of a recognition-mediated duplex. Design of the self-complementary motif, necessary for the internal, template-mediated stabilisation was inspired by Rebek's replicator,[109] introduced previously. Condensation of three aldehydes (**Al^1** to **Al^3**) and two amines (**Am^1** and **Am^2**), afforded six different imine products (**Figure 1.36**). Only the imine product **Al^1–Am^1** was capable of forming a recognition-mediated duplex [**Al^1–Am^1**·**Al^1–Am^1**].

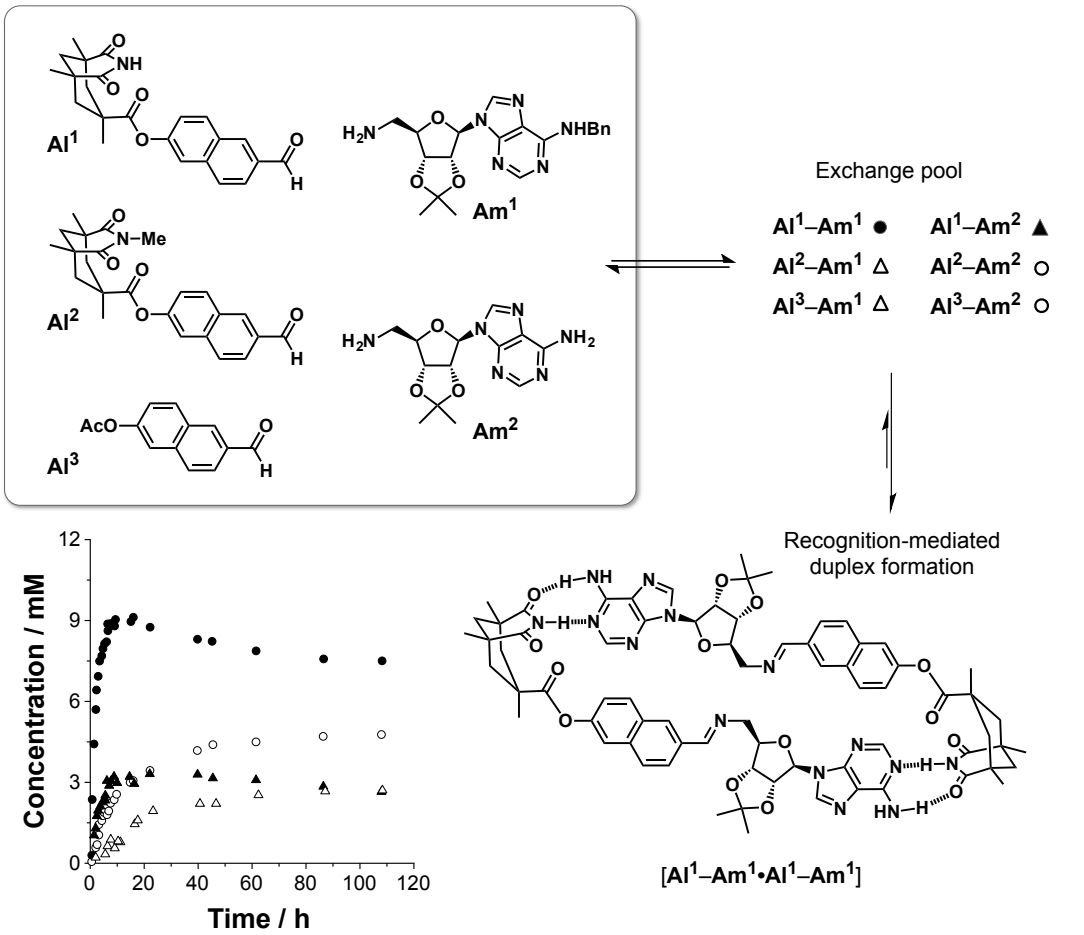


Figure 1.36 Dynamic library composed from three aldehydes, **Al^1** to **Al^3** and two amines, **Am^1** and **Am^2**, affording 6 different imine products. Only imine **Al^1–Am^1** (black circles) is capable of stabilisation through the formation a recognition-mediated duplex [**Al^1–Am^1**·**Al^1–Am^1**]. The initial advantage afforded in rate eroded over time as **Al^2–Am^2** and **Al^3–Am^2** (empty circles) products formed as the more thermodynamically stable products. The imine **Al^1–Am^2** (black triangle) product was similarly affected, decreasing in concentration over time. **Al^2–Am^1** and **Al^3–Am^1** (empty triangles) formed slowly and unselectively. Figure adapted from Ref. 147.

Kinetic analysis of the library revealed that the initial advantage in the rate of formation of **Al[1]–Am[1]**, afforded by the recognition mediated binding, eroded over time. **Figure 1.36** (black circles) shows that this erosion stems from the thermodynamic preference of the library for products **Al[2]–Am[2]** and **Al[3]–Am[2]** (empty circles). The study showed clearly that the level of selectivity that can be achieved for a product formed *via* recognition-mediated reaction processes in the library is limited by the thermodynamics stabilities of the imine products, and the reversible nature of the system.

In order to investigate the complexities associated with the formation of a replicator based on reversible bond formation, the Philp laboratory developed[148] an imine-based system (**Figure 1.37**) formed from an aromatic aldehyde **47** and an amine **48**. These two components can react to form imine **49**, which is in dynamic equilibrium with its constituents. Owing to the complementary nature of the carboxylic acid and dimethyl amidopyridine unit, the formed imine has the capacity to assemble the unreacted aldehyde and amine in a catalytically active ternary complex [**47**·**48**·**49**] that can accelerate their reaction. Interestingly, the authors showed that while the addition of preformed template of imine **49** removes the lag period observed in the absence of template, this addition also results in a decrease in the overall quantity of the newly formed imine **49** produced within the system.

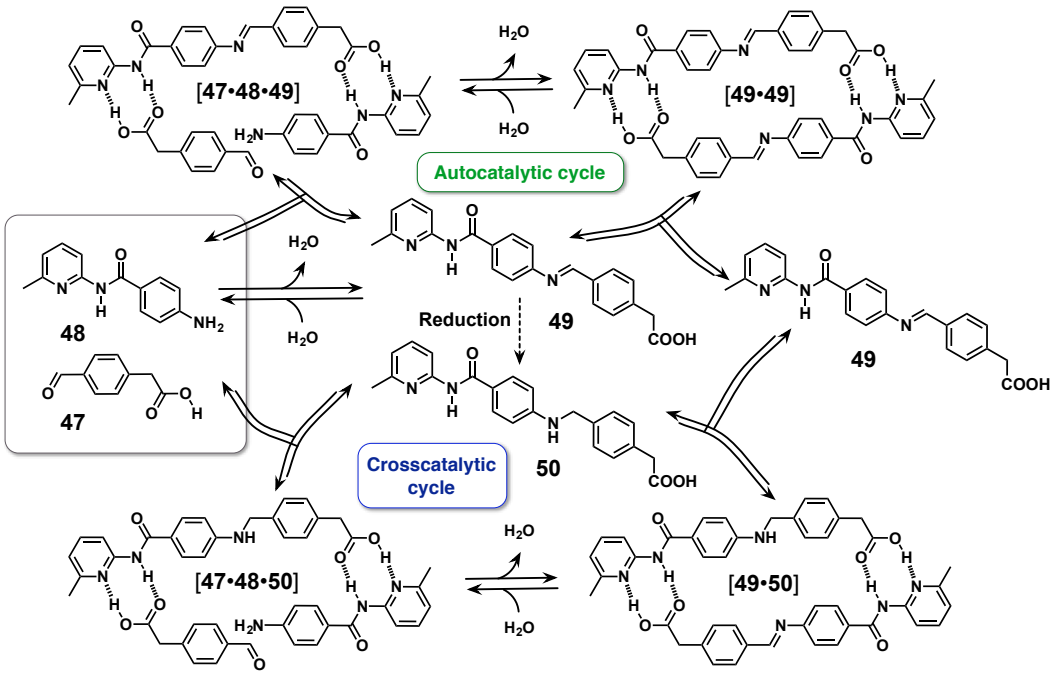


Figure 1.37 A dynamic imine system incorporating formation of a self-replicator. Aldehyde **47** and amine **48** can react reversibly to form imine **49**, which is capable of templating its own formation *via* the catalytically-active ternary complex [**47**·**48**·**49**]. The reversible nature of the covalent bond forming step employed, however, imposes a limit on the production of the self-replicating imine **49**. Reduction of the imine to the corresponding amine **50** afforded a crosscatalytic template. Figure adapted from Ref. 148.

In fact, kinetic simulations revealed that the decrease is proportional to the amount of template added—a thermodynamic boundary imposes a limit on the equilibrium position for the formation of this imine, and addition of preformed template is not sufficient to break away from this position. Instruction of this dynamic imine system with a reduced amine counterpart (**50**) of the imine **49** confirmed that this crosscatalytic pathway operates efficiently, and without the decrease in imine formation observed previously in the presence of **49**. Nevertheless, even the addition of the reduced template **50** did not allow the system to shift away from its thermodynamically most stable position.

1.5.2 Limits on selection and amplification in DCLs under thermodynamic control

DCC has been exploited in numerous examples over the past twenty years—and the majority of these focus, as illustrated by the few selected examples in the previous section, on dynamic systems under thermodynamic control. The thermodynamic nature of the DCLs, however, presents limitations on the level of amplification that can be achieved upon addition of a stimulus capable of binding one or several of the library components. In fact, fairly few of the DCC systems show high level of amplification for a single product. Numerous research groups have explored the theoretical limitations on selectivity in libraries under thermodynamic control, and selected results will be reviewed briefly. In 2000, Moore and Zimmerman reported[149] a critical theoretical study examining whether molecular recognition can be used to drive re-equilibration of a library comprised of a large population of copolymer sequences, capable of interchanging dynamically, to form a larger subset of sequences with a high affinity for the added ligand (so-called masterpiece sequences) (**Figure 1.38**). The model simulated a dynamically-exchanging population of copolymer sequences, each capable of binding a ligand with a specific binding constant (K). The binding constants for the library of copolymers are assigned using a normal probability distribution function (in log K), with a mean $\mu = 0$ and standard deviation $\sigma = 1$. In the absence of ligand, however, each component is present at an identical concentration.

Behaviour of the library of copolymers was simulated in the presence of increasing amounts of ligand, showing progressive shift in the mean binding constant distribution curve towards higher mean K. However, the maximum increase in the mean binding affinity, obtained in the presence of excess ligand, was only *ca.* two orders of magnitude larger than in the absence of any ligand. This model shows that a small proportion of masterpiece sequences, capable of binding the ligand extremely strongly, is not capable of shifting the library mean binding constant by more than two degrees of magnitude.

Instead, the magnitude of re-equilibration is countered the presence of a large number of weakly binding sequences. Therefore, in order for one or more strongly binding components to be amplified from a large library, methods that would permit the system to break away from its equilibrium position need to be employed.

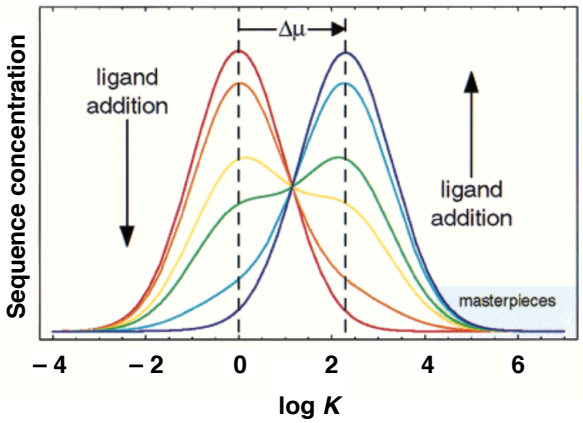


Figure 1.38 Simulation outcome of a population of interchangeable copolymer sequences, with normally distributed binding affinities across the population (mean, $\mu = 0$; st. deviation, $\sigma = 1$) in the presence of increasing concentrations of ligand. The model shows that the mean (geometric) binding constant can be shifted by two degrees of magnitude (log K). Figure adapted with permission from Ref. 149. Copyright 2000 American Chemical Society.

While providing information about the bulk behaviour of the library, the Moore and Zimmerman study attracted criticism[150] by Otto, Sanders and co-workers for having simulating the library as a continuum of interchanging species. Specifically, these authors argued[150] that real interest in utilising DCLs for discovery of high affinity binders is to be able to identify the masterpiece species with the strongest binding, rather than the behaviour of the library as a whole. These authors have extended[150] the previously reported simulation (**Figure 1.39**), examining the library population as a collection of discrete species (10000 members), each with a randomly assigned binding constant K (the same mean and standard deviation as employed in the Zimmerman study). As in the Moore and Zimmerman system, the library components maintained the capacity to interchange, and, in the absence of added ligand, were present at an identical concentration. The authors investigated how the concentration of the best binders changes after addition of template.

The simulated DCL in the presence of excess template ([template] = 10 M, [combined library members] = 1 M) showed the general behaviour and distribution trends reported by Moore and Zimmerman, particularly at low binding constants. On the strong extreme of binding, however, the distribution determined by Otto, Sanders and co-workers shifted to a single library component, amplified $\times$ 794, comprising 8% of the entire library. The authors have also undertaken the same simulation, varying

the number of library components from 10 to 10^6, showing that the mean yield of the highest affinity binder decreases fairly slowly with an increasing library size. For example, for a DCL built from 1000 components, the best binder will comprise *ca.* 13% of the library, whereas for the largest size DCL simulated (10^6 members), the proportion of the best binder falls to 0.5%. The authors argued that the slow decrease results from the interplay of two factors. Firstly, the larger the library, the higher the chance of finding a binder capable of stronger affinity. Contrastingly, the larger library size equates lower initial concentration, and, thus, also a lower amplified concentration. The authors aptly noted that while these yields might seem low compared to those obtainable by traditional synthetic methods, the primary utility of DCC is in its capacity to allow efficient screening of a large number of compounds. Once identified, the best binders can be examined in isolation in order to provide the desired species in much higher yields.

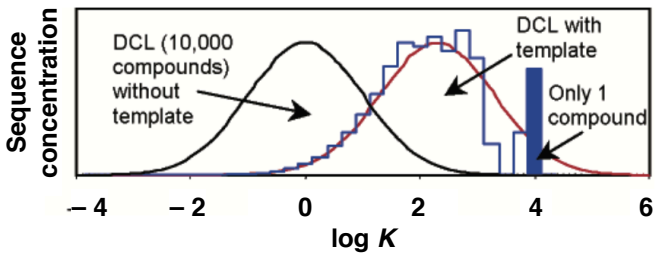
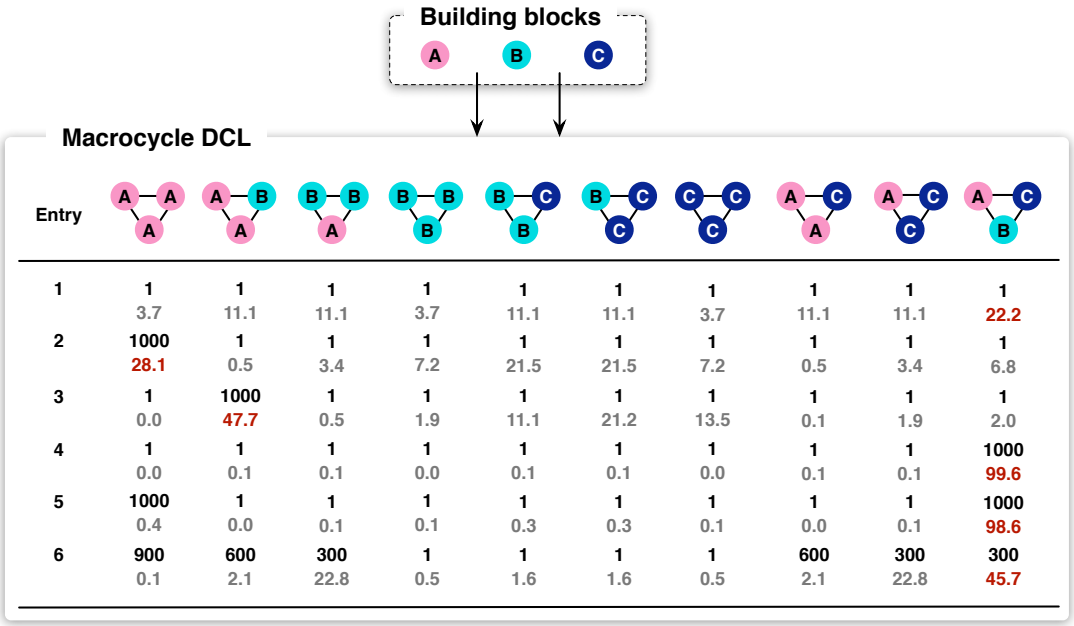


Figure 1.39 Histograms showing the distribution of a continuous DCL (10000 members) as reported by Moore and Zimmerman in the absence (black) and in the presence (red) of a template. Typical histogram of a simulated DCL (blue) instructed with ligand. Bar height reflects the overall concentration of the species with that particular *K* log in the library. The best binder (single compound with *K* log = 4) makes up 8% of the library. Figure reprinted with permission from Ref. 150. Copyright 2004 American Chemical Society.

These two theoretical studies examining amplification in DCLs have clear benefits, affording a facile route to exploring a large number of DCLs, particularly libraries with a very large number of components, the analysis of which could prove difficult in practice. Nevertheless, both reports are limited by the assumptions employed. Namely, the ability of each library to interchange in a 1:1 fashion, and the equal concentration of each library component at the onset of the simulation. Severin and co-workers have reported[151] simulations of a more realistic library, comprised of three different building blocks, **A**, **B** and **C**, capable of reversibly forming 10 different macrocyclic assemblies with a fixed stoichiometry (X_3). Through simulations, the authors showed (**Figure 1.40**) that this type of library is fundamentally different in the displayed behaviour, compared to a library where a single building block is capable of assembling into products with variable stoichiometry.



Entry	A–A (A)	A–B (A)	B–B (A)	B–B (B)	B–C (B)	B–C (C)	C–C (C)	A–C (A)	A–C (C)	A–C (B)
1	1 3.7	1 11.1	1 11.1	1 3.7	1 11.1	1 11.1	1 3.7	1 11.1	1 11.1	1 22.2
2	1000 28.1	1 0.5	1 3.4	1 7.2	1 21.5	1 21.5	1 7.2	1 0.5	1 3.4	1 6.8
3	1 0.0	1000 47.7	1 0.5	1 1.9	1 11.1	1 21.2	1 13.5	1 0.1	1 1.9	1 2.0
4	1 0.0	1 0.1	1 0.1	1 0.0	1 0.1	1 0.1	1 0.0	1 0.1	1 0.1	1000 99.6
5	1000 0.4	1 0.0	1 0.1	1 0.1	1 0.3	1 0.3	1 0.1	1 0.0	1 0.1	1000 98.6
6	900 0.1	600 2.1	300 22.8	1 0.5	1 1.6	1 1.6	1 0.5	600 2.1	300 22.8	300 45.7

Figure 1.40 Simulated steady-state concentrations of a dynamic mixture of macrocycles with the fixed stoichiometry X_3 obtained by assembly of three building blocks **A**, **B** and **C** (log $K_{CCC} \gg 5$; $[\mathbf{A}]_{total} = [\mathbf{B}]_{total} = [\mathbf{C}]_{total}$). Numbers in black represent the relative stabilisation factor while grey numbers show the relative concentration of each macrocycle (100% overall). Text in red highlights the percentage of macrocycle present at the highest concentration within the mixture in each condition. Figure adapted from Ref. 151.

In the absence of thermodynamic preferences (absence of template), a statistical mixture of macrocycles will be obtained (**Figure 1.40**, entry 1). Entries 2 to 4 examine the library under conditions where a single member of the library is stabilised (by a factor of 1000) by binding to a guest molecule. The most dramatic amplification is observed when the heterotrimeric macrocycle **ABC** is stabilised, revealing that this species is formed almost exclusively. This behaviour is in stark contrast with the stabilisation of **AAA** or **AAB**. In both cases, the amplified product dominates the library composition, however, species incorporating **B** and **C** building blocks are also amplified. In a scenario (entry 5) where both **AAA** and **ABC** are stabilised simultaneously, remarkably, only **ABC** is amplified significantly. **AAA** on the other hand, is virtually extinct in the library. Even in a situation with more distributed amplification (entry 6), the heterotrimer **ABC** remains the dominant species.

The simulation results show that in a DCL of the type investigated here, it is not always the most stable (stabilised) library member that is amplified the most. A species can become extinct despite being having the capacity to form the most stable product (entry 6) and, instead, the DCL members that reduce the overall free energy of the library the most are amplified. Severin and co-workers have demonstrated[151] the principles

determined using DCL simulations also experimentally, utilising dynamic mixtures of metallomacrocycles. Building on previously reported work on the assembly of trinuclear complexes using organometallic pyridonate systems, the authors exploited steric effects as a tool to induce thermodynamic stability variation amongst library members. In the experimental system, the most stable (the least sterically disfavoured) homotrimeric assembly was never found to be the preferred species, and, instead, amplification of the heteromeric assemblies was preferred. The preference for the formation of the heteromeric assemblies stems from the fact that it is more energetically favourable for the library to produce, from a set of three building block, a larger number of, albeit potentially weaker, complexes as opposed to a smaller number of very strong assemblies. This preference is supplemented by the fact that even in the absence of template, the heteromeric assembly is preferred statistically (**Figure 1.40**, entry 1). For a similar reason, in a dynamic library that can form macrocycles with various stoichiometries (*e.g.* X_3 or X_5), a larger number of smaller macrocycles is more favourable to a smaller quantity of larger macrocycles.

In further work on DCL libraries, Severin has explored[152] target-induced adaptation and selection using numerical simulations. The results agreed with the previously reported observation that the highest affinity binder is often not the one that is amplified the most. In fact, the presence of the target might result in a decrease in the concentration of the best binder at equilibrium. This finding is essential to understanding of this type of DCLs. In order to achieve high selectivity in the library component capable of binding a specific target the most strongly, and, thus, to utilise DCL as a selection strategy for the identification and isolation of a component with the most favourable properties, relies on the best binder being the most amplified species. Using numeric simulations, Severin also identified[152] numerous approaches that can be employed in order to make the strongest binder the component made preferentially. These methods include using substoichiometric amounts of a shared building block and low concentrations of target template. Interestingly, Severin also proposed the idea of utilising the so-called evolutionary protocols (**Figure 1.41**) that exploit immobilised target and repeated rounds of equilibration, change of conditions to freeze the equilibrated library and separation of bound DCL members from those not bound to the immobilised target.

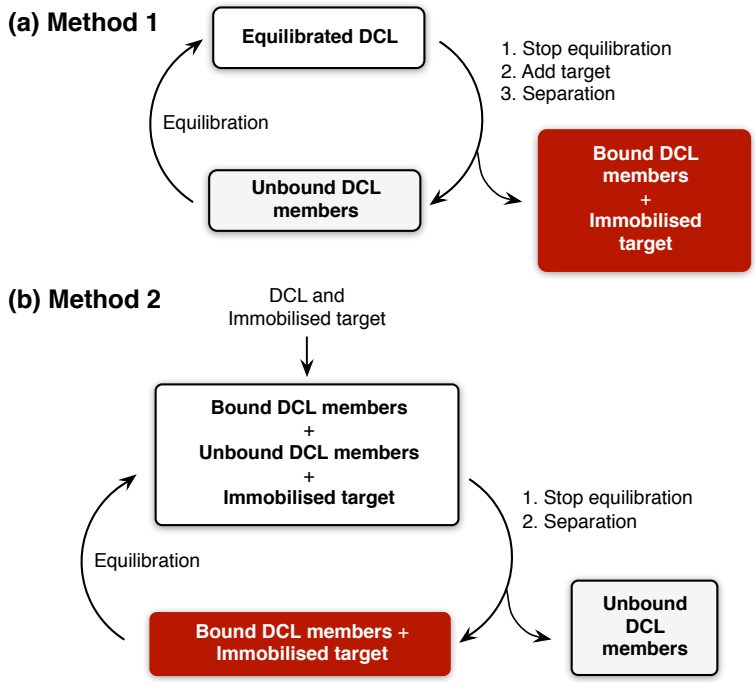


Figure 1.41 Two alternative methods, exploiting iterative procedures for the selection of DCL members exhibiting high affinity for an immobilised target as proposed by Severin. Figure adapted from Ref. 152.

These findings reported by Severin are supported by the work of Sanders and co-workers, who investigated[153] the correlation between host-guest binding and the subsequent amplification in simulated dynamic libraries. Investigating over 50 libraries, with component number from 65 to 4828, the authors focused on determining the affect of template and library concentration on the said correlation. The authors found that the competition of library products for building blocks is the factor preventing amplification of the highest-affinity binder. In general, the simulations showed that employing low library concentrations and substoichiometric amounts of template correlate with high likelihood of the best binder achieving high amplification. Conversely, excess of template, *i.e.* a situation where a sufficient amount of template is present for binding with more than the best binding components, showed low correlation between the strength of binding and the degree of amplification.

More recently, this work was ellaborated[154] by Ludlow and Otto, who explored the effect of library size on the likelihood of detecting (i) any amplification at all, (ii) the highest affinity library component and (iii) binding affinity of the most amplified DCL component. The study was directed at examining the question of whether the strongest binders are identified best using a single large library, where there is a higher chance that the DCL contains a better binder but simultaneously detection of all library members might be more difficult, or multiple smaller libraries. Within the simulated

libraries, binding affinities were assigned randomly according to a normal distribution function and detection limits were estimated using the experimental limit of LC-MS analyses. The simulations of various library sizes, systematically varying building block and template concentrations showed that ultimately, larger libraries produce stronger binders. Specifically, this outcome stems from the interplay of two factors: the probability of finding a strong binder increases more rapidly with library size than the detection limit drops. Moreover, larger DCLs showed a lower likelihood of exhibiting no amplification whatsoever. The authors also argued that in order to achieve the most favourable trade-off in obtaining a high-affinity ligand, while being still able to detect it, a 10:1 ratio of library building blocks to template is the most favourable in the investigated systems.

The theoretical studies discussed in this section, together with further work on selectivity in DCLs by Reek[155] and Sanders,[156] address fundamental questions regarding the limits on selectivity and amplification in dynamic covalent libraries—the aspects at the very core of dynamic covalent chemistry. The outcomes reveal the efficiency and utility of theoretical simulations, often based on experimental systems, in ascertaining the effect of various initial conditions, as well as the library size and thermodynamic parameters on the adaptive behaviour of dynamic libraries of interchanging components. Importantly, the theoretical work has the capacity to address the gap in experimental work, arising from analytical difficulties and time limitations. The presented works show that it is possible to fine-tune the initial library conditions, such as the template and building block concentration and number in order to increase the likelihood of detecting the high-affinity binders, as well to increase the magnitude of the binding affinity itself.

1.5.3 Dynamic covalent systems coupled to kinetically-driven irreversible reaction processes

The selection strategies, and, thus also the resulting degree of amplification, described thus far correlate with the amount of template added and the concentrations and binding affinities of the components making up the library. In addition to fine-tuning these parameters as a strategy for obtaining high selectivity for the best affinity binder, a dynamically-exchanging covalent system can be coupled[134,157,158] to kinetically-controlled processes. In this way, the limits imposed on the system by the thermodynamic regime can be broken. By exploiting methods that allow the added target *e.g.* template) not only to interact but also to react with the library components, the best binders can be irreversibly removed *via* kinetic selection processes from the thermody-

namic pool. Additional benefit arises from utilising kinetically-controlled processes, *i.e.* easier isolation of the amplified, now more stable library products. As well as potentially enhanced selectivity, the combination of thermodynamic and kinetic processes within a single system presents scientist with the possibility of exploring networks and systems with an additional layer of complexity—more similar to the varied, large-size interconnected systems found in nature.

An interesting example of a dynamic covalent library driven by kinetic selection was reported[159] in 2002 by Kazlauskas *et al*. The method and experimental setup, termed[136] by the authors as a pseudo-dynamic library, was comprised of a pool of different dipeptide inhibitors of carbonic anhydrase (CA), which was present as the receptor within one reaction chamber of the library (**Figure 1.42**). In another chamber, separated by a dialysis membrane, a protease enzyme (pronase) was present, capable of destroying the dipeptide inhibitors. Therefore, the dipeptides capable of forming the strongest complex with the receptor were sequestered within the cavity of the CA in one chamber, away from the protease. The remaining unbound dipeptides could diffuse through the membrane into a chamber containing the pronase, resulting in their hydrolysis to give the corresponding amino acid residues.

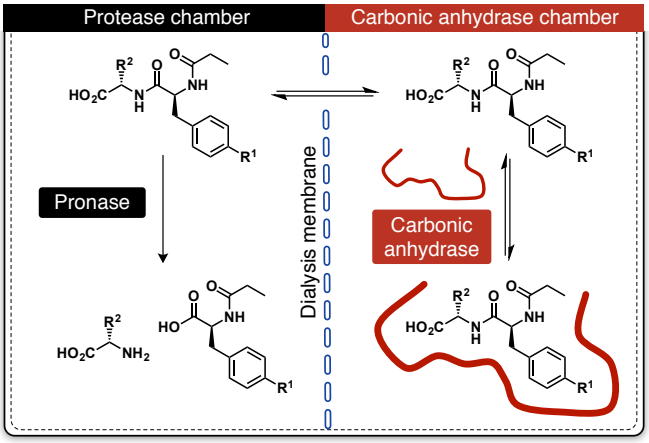


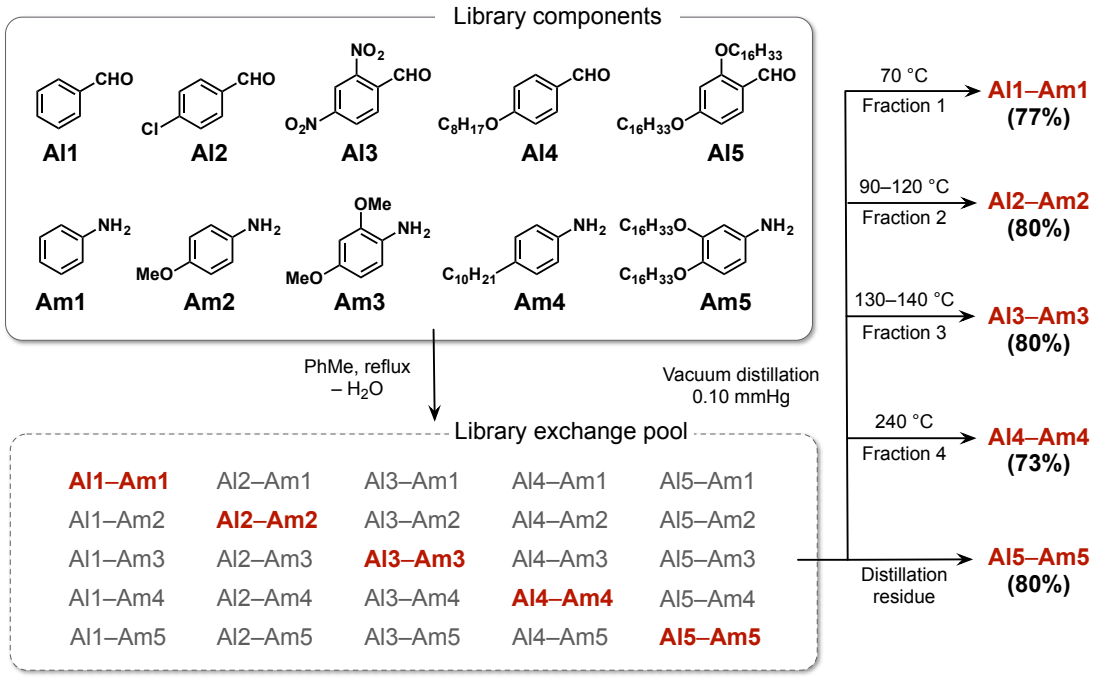
Figure 1.42 Experimental setup of a pseudo-DCL exploiting kinetically controlled destruction. A library of dipeptide inhibitors of carbonic anhydrase can either bind the anhydrase within one chamber or diffuse through a dialysis membrane into a chamber containing protease, which cleaves the dipeptide into the corresponding amino acids. Figure adapted from Ref. 159.

The authors examined the selectivity for a well-established dipeptide inhibitor with a high affinity for CA relative to one with weak binding properties. The results after 12 hours showed a significantly higher selectivity for the established CA inhibitor than would be predicted solely based on the thermodynamic boundaries governing the system. While this system is perhaps not easily adaptable for small molecule DCL

alternatives, the pseudo-DCL method illustrates the possibility of breaking away from the thermodynamic barrier through methods of kinetically removing DCL components.

The DCC approach exploited by Kazlauskas and co-workers, while successful in amplifying a product selectivity, could be considered to be fairly destructive. The authors have expanded[160] the reported pseudo-DCL model to include a third synthesis chamber, where the hydrolysed amino acids were employed in dipeptide synthesis, thereby amplifying further the selectivity for the best binders. Nevertheless, the concept of incorporating kinetically-controlled processes as a selection strategy can be effective also in systems utilising constructive methods. The kinetic means of directing a dynamic library to select a single product, and, thus, remove the component irreversibly from the library can be achieved using physical and chemical processes, or their combination. While focus in this introduction will be on libraries coupled to chemical reactions, illustrative examples of both approaches will be described and discussed in the following section.

Milanjić and co-workers have exploited[161,162] distillation in order to achieve selectivity, as well as mechanical separation of products from a dynamic library of imines. In the earlier work, the authors explored[161] library of interconverting imines, assembled from aromatic aldehydes and anilines, affording a library of imines, each with a specific boiling point, determined by the combination of its constituent elements. The authors achieved selective amplification by exploiting the reversible nature of the imine bond and the Le Châtelier principle—in a library where exchange reactions operate faster than the rate of selective removal by distillation, extraction of the lowest boiling point library component disturbs the equilibrium, forcing the other components to re-equilibrate so as to produce more of the component being extracted. After a product and its constituents are extracted from the reaction mixture, the process can be repeated to obtain multiple species selectively, selected on the basis of their volatility. The authors demonstrated the principle on progressively larger libraries, ranging from the smallest (two aldehydes and two imines) to the most complex, comprised of five aldehydes, **Al1** to **Al5** and five amines, **Am1** to **Am5** (**Scheme 1.3**). While the formation of all exchange pool products was not confirmed, a library comprised of up to 25 imines was successfully resolved using distillation to produce five different compounds (**Scheme 1.3**, red), all in yields above 70%.



Scheme 1.3 Dynamic covalent imine library assembled from five aldehydes, **Al1** to **Al5** and five amines, **Am1** to **Am5**. The resulting virtual library was physically transformed by the means of distillation, producing five different compounds (highlighted in red) selectively, in yield > 70%. Scheme adapted from Ref. 161.

More recently, the authors have applied[162] the same principle to resolution of dynamic ester libraries. Using metal alkoxides such as NaO*t*-Bu or $Ti(OBu)_4$ to initiate acyl exchange, the authors employed successfully distillation of an ester library assembled from up to four carboxylic acids and four alcohols, obtaining four industrially relevant esters (*e.g.* ethyl acetate and butyl butyrate) in a single reaction setup.

Blackmond and co-workers utilised[163] crystallisation process as a means of resolving a dynamic equilibrium between two imines, (*S*)-**51** and (*R*)-**51**. Imine **51** racemises in solution to give the *R* and *S* species. The authors examined imine **51** in a dynamic equilibrium between the solid and solution phase in MeOH or MeCN, in the presence of DBU (1,8-diazabicyclo[5.4.0]undec-7-ene) as a non-nucleophilic base (**Figure 1.43a**). The enantiomeric excess (ee) of the dynamic system was found to evolve over time in the presence of glass beads as a source of mechanical energy—from an initial ee imbalance in crystal composition as low as 3% (**Figure 1.43b**) to a single chiral solid state (*R* or *S*).

A crystallisation-like process was also exploited by the laboratories of Otto and Ashkenasy, for the development of fibre[164–166] and β-sheet[167,168] based replicating systems, respectively. In these larger systems, the mode of replication is significantly less well-defined than in systems based on small synthetic molecules or oligonucleotides,

relying on agitation to produce free fibre ends, which nucleate the growth of fibres or β-sheets.

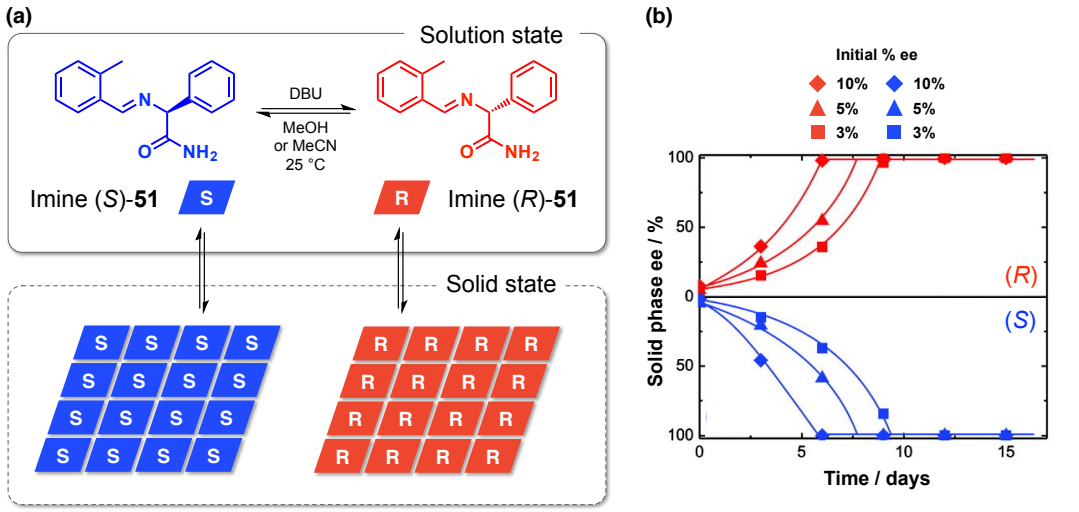
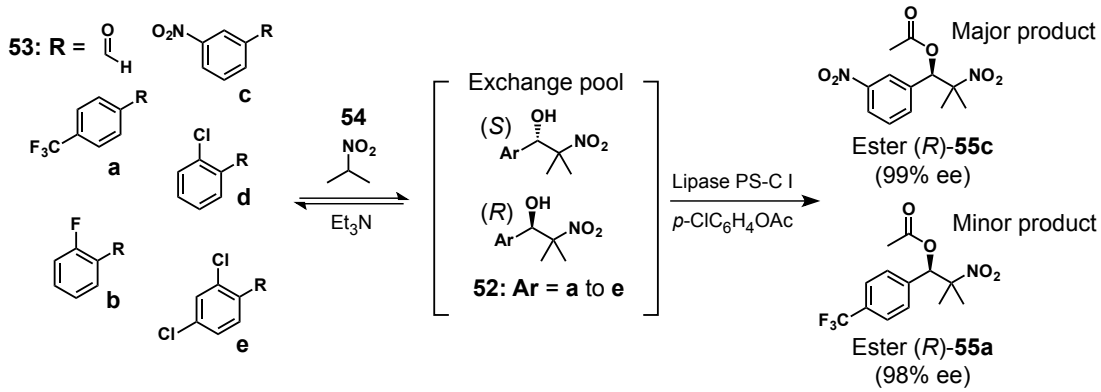


Figure 1.43 **(a)** In basic MeOH or MeCN solutions, imine **51** equilibrates between (*R*)-**51** and (*S*)-**51** states. **(b)** In the presence of glass beads, a basic MeCN solution containing solid phase (*R*)-**51** or (*S*)-**51** at various initial ee (enantiomeric excess) evolves towards a single chiral solid state over time. Figure adapted from Ref. 163.

Ramström and co-workers have developed[169,170] a lipase-mediated resolution of a dynamic library. In this work, library of 10 racemic nitroaldols **52a** to **e**, is formed through the reversible reaction of five aromatic aldehydes equipped with various substituents, **53a** to **53e**, and 2-nitropropane **54** in the presence of Et_3N (**Scheme 1.4**).

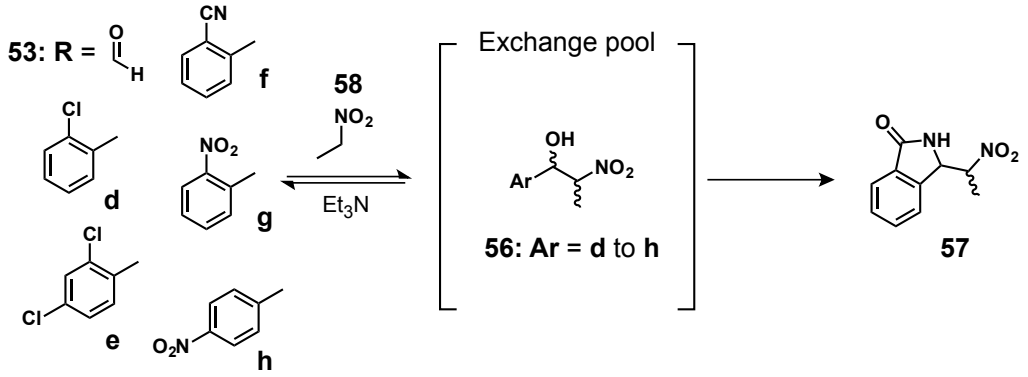


Scheme 1.4 A dynamic racemic nitroaldol library assembles from five aldehydes, **53a** to **53e** and 2-nitropropane **54** under basic conditions. Addition of lipase (lipase PS-C I) and an acyl donor (*p*-ClC_6H_4OAc) results in dynamic kinetic resolution of the library of nitroaldols **52a** to **e** to produce two ester products, (*R*)-**55c** and *R*)-**55a**, selectively. Scheme adapted from Ref. 169.

Upon addition of a lipase (an esterase) PS-C I and *p*-chlorophenyl acetate as the acyl donor to the library at 40 °C resulted in selective production of two esters, (*R*)-**55a**

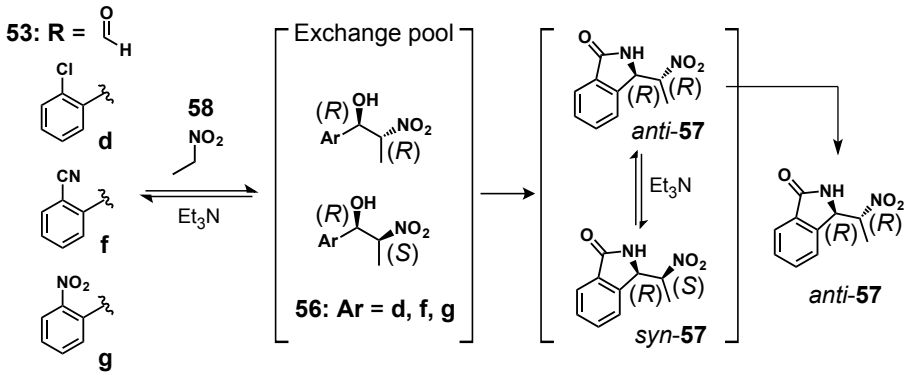
and (*R*)-**55c** as the major product and (*R*)-**53a** as the minor component. The lipase-mediated reaction amplified (*R*)-**55c** successfully from the library mixture, despite the concentration of the corresponding nitroaldol component being amongst the lowest in the library. The authors found that prolonged reaction time enabled isolation of the two lipase-products in increased yields (24% after 24 hours versus 80% after 14 days).

In related work, Ramström and co-workers exploited[171] a tandem reaction to drive the resolution of a nitroaldol library. The library was again comprised by five benzaldehyde derivatives, **53d** to **53h** (**Scheme 1.5**). In this library, aldehyde **53f** only possesses a 2-nitrile substitution pattern required for cyclisation of the nitroaldol exchange pool product **56** to an iminolactone **57**. Mixing of these five aldehydes with nitroethane **58** in acetonitrile allowed an exchange pool of nitroalcohols to form (**Scheme 1.5**). Already after 30 minutes, however, formation of a single new product began to dominate the reaction mixture—iminolactone **57**. After 24 hours, all nitroalcohols and nitroethane **58** present in the library have been consumed in favour of **57**.



Scheme 1.5 A dynamic nitroaldol library assembles from five aldehydes, **53d** to **h** and nitroethane **58** under basic conditions. Resolution of the nitroaldol library (**56d** to **h**) is driven by cyclisation of nitroaldol component incorporating 2-nitrile substituted benzaldehyde **53f** to produce iminolactone **57** selectively. Scheme adapted from Ref. 171.

The stereochemistry of this nitroaldol system was explored[172] by Ramström and co-workers (**Scheme 1.6**). The authors focused their attention on three aldehydes in particular, **53d**, **53f** and **53g**. In the presence of nitroethane **58**, these aldehydes equilibrated with the nitroaldol products **56d**, **56f** and **56g**. Incorporation of only one aldehyde (**53f**) enabled the resulting nitroaldol **56f** to cyclise irreversibly, producing intermediate iminolactone **57**. In basic solution, this product is formed in two diastereoisomeric forms, *anti* and *syn*, in equilibrium with one another. One of the diastereoisomers exhibited a significantly higher propensity for crystallisation, a fact that was exploited by the authors in order to achieve diastereomeric amplification driven by phase-change. In this particular system, the authors were able to identify the amplified species as the *anti*-**57** (*R*,*R*)/(*S*,*S*) diastereoisomer (**Scheme 1.6**).



Scheme 1.6 A dynamic nitroaldol library assembles from three aldehydes, **53d**, **53f** and **53g** and nitroethane **58** under basic conditions. Resolution of the nitroaldol library (**56d**, **56f** and **56g**) is driven first by cyclisation of nitroaldol component incorporating 2-nitrile substituted benzaldehyde **53f** to produce the two diastereoisomeric iminolactones *syn*-**57** and *anti*-**57** selectively. Secondary selection in the library is driven by phase change, which allows selective crystallisation of the *anti*-**57** (*R*,*R*)/(*S*,*S*) diastereoisomer. For clarity, only one enantiomeric form is shown for each diastereoisomers. Scheme adapted from Ref. 172.

These presented systems were expanded and altered by Ramström and co-workers further to produce lipase-driven resolution (i) of a doubly dynamic imine-Strecker reaction (cyanation) system[173] (ii) resolution of a dynamic cyanohydrin system,[174] (iii) resolution of a dynamic mixed nitroaldol–thioester system[175] and (iv) lipase-catalysed selective lactonisation of a dynamic hemithioacetal system.[176] These examples illustrate the utility[177–179] of irreversible chemical and physical processes, often employed in tandem, as tools in selective resolution of dynamic covalent systems.

The lipase-mediated resolution of dynamic covalent systems is a recognition-mediated process, where species which are effectively the best substrates for this enzyme are amplified and transformed irreversibly. The lipase alters these suitable substrates, removing them from the interconverting dynamic pool—an analogous means of amplification can be envisaged in a dynamic system where the added target not only recognises an exchange pool component, but also has the capacity to react with it irreversibly (**Figure 1.44**, AB). The degree of perturbation to a dynamic system, and, thus, also the level of target amplification, can be magnified even further if the irreversible reaction was capable of feeding back into the reaction, for example by template-mediated autocatalysis (**Figure 1.44**, SR). In this case, an exchange pool of interconverting components (**Figure 1.44**) contains a few selected members with the capacity to interact and react with an added target. However, only a small subset of these components has the ability to both react and interact, resulting in amplification *via* irreversible recognition-mediated or template-mediated processes.

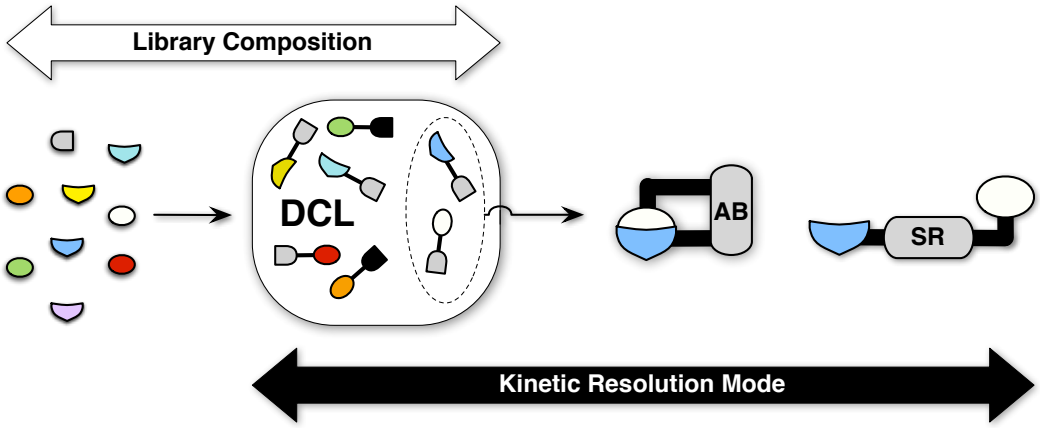


Figure 1.44 Resolution and thus also final distribution of a dynamic covalent library (DCL) depends on the interconverting components comprising the library and the method of resolution employed. In the cartoon representation shown here, two irreversible kinetically-control recognition-mediated means of library resolution are presented: AB mode of resolution denotes a method driven by reaction of selected library members in a recognition-mediated reaction operating through a binary reactive complex. SR represents a resolution method driven by reaction of library members in a template-mediated autocatalytic reaction proceeding through a ternary catalytically-active complex.

Having demonstrated[143] that nitrones exchange dynamically in non polar solvents such as $CDCl_3$, Philp and Sadownik examined[180] the effect of coupling a dynamic covalent imine-nitrone system to a recognition-mediated irreversible reaction. Specifically, the dynamic library is built from two aldehydes, **59** and **60**, 4-fluoroaniline **61** and 4-fluorophenylhydorxylamine **35**. The library (**Figure 1.45**) exchange pool thus contains two imine products, **62** and **63** and two nitrones, **64** and **65**. Only the two nitrone components possess the reactive site required for 1,3-dipolar cycloaddition with a maleimide. The maleimide **66** used to irreversibly transform selected components was designed to contain a carboxylic acid recognition site. Therefore, reaction of the library component bearing a complementary amidopyridine recognition site, nitrone **65**, is not only irreversible but also recognition-mediated. The nitrone component **65** can associate with the maleimide **66** in a binary reactive complex [**65**·**66**] that accelerates the reaction between these two components, leading to selective amplification of the *cis* diastereoisomer *cis*-**67** (**Figure 1.45**, red box). Nitrone **64**, lacking the recognition site, can react with the same maleimide only *via* the slow bimolecular reaction channel, producing *cis*-**68** and *trans*-**68** products with low diastereoselectivity.

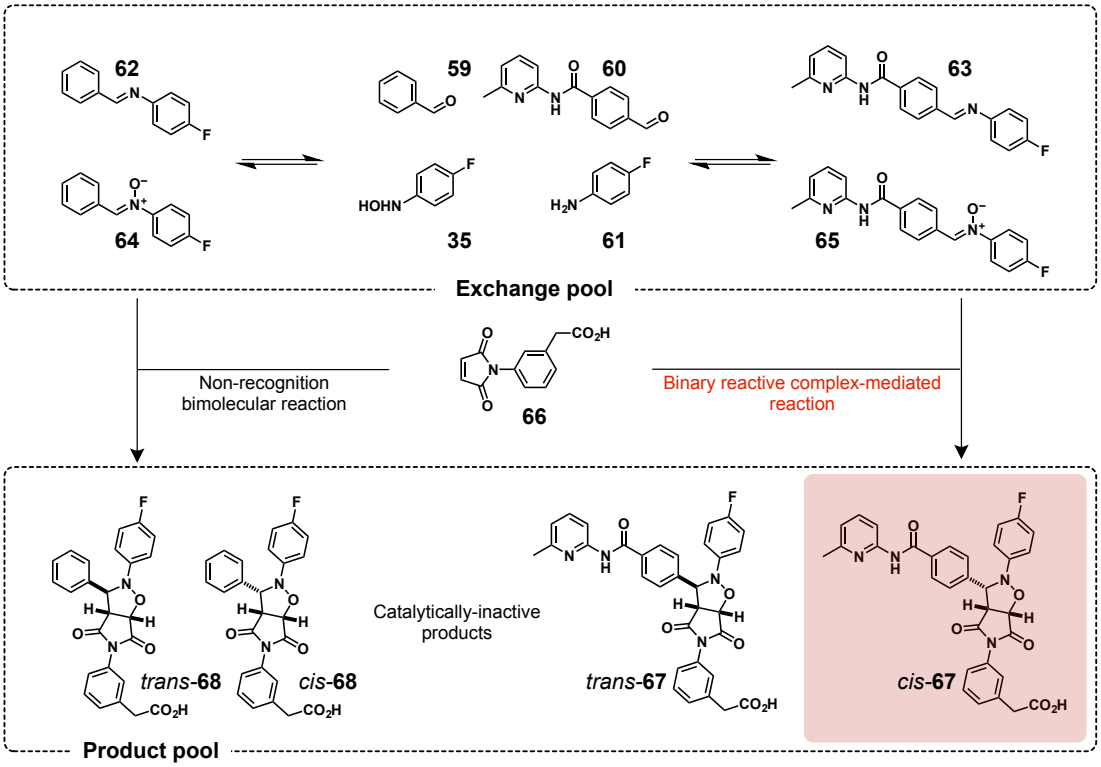


Figure 1.45 A dynamic library comprised of two interconverting imines, **62** and **63**, and two nitrone components, **64** and **65**. Within the library, two components are equipped with a reactive nitrone site and two with an amidopyridine recognition site. However, only a single component, nitrone **65**, possess both the recognition and reactive elements required for the irreversible reaction with maleimide **66** *via* the reactive binary complex between **65** and **66** to produce *cis*-**67** selectively. Nitrone **64** reacts with maleimide **66** only through a slow bimolecular reaction, producing both *trans*-**68** and *cis*-**68**, with low diastereoselectivity. Figure adapted from Ref. 180.

Initially, the authors verified the dynamic nature of the system and its equilibrium position by mixing imine **62** and nitrone **64** in CD_2Cl_2 saturated with *para*-toluene sulfonic acid monohydrate (*p*TSA). Within hours, the library equilibrated to give a mixture of all imines and nitrones. Repeating the process with imine **63** and nitrone **65** resulted in essentially the same equilibrium position. The authors found that a library initially assembled from imine **62** and nitrone **64** is transformed irreversibly in the presence of maleimide **66**—maleimide conversion reached 72% after 16 hours, and the cycloadduct *cis*-**67** is formed rapidly and with high diastereoselectivity, constituting more than 90% of the library product pool. In this manner, the authors were able to illustrate that even a single, carefully-designed recognition-mediated irreversible process can direct the library away from its original equilibrium position, transforming the library to produce a single product with high selectivity.

The dynamic system resolved by the means of a single recognition-mediated irreversible process was expanded by Philp and co-workers[181] to a larger, 24 component

dynamic covalent library (**Figure 1.46**) coupled to two recognition-mediated reaction processes. In addition to the aldehyde **60**, amine **61** and hydroxylamine **35** employed previously, the library also contained three additional aldehydes (**69** to **71**) and two anilines (**72** and **73**). The additional aldehyde **69**, and, thus also the nitrone **74** is equipped with a 4,6-dimethylamidopyridine recognition unit that is also capable of recognising the carboxylic acid on maleimide **66**, and, therefore, two competing recognition-mediated reaction processes can operate in the library. The authors investigated this library with a view to examining the selectivity for these recognition-mediated reaction processes.

The authors were particularly interested in examining how the selectivity within a DCL for these processes compares to that obtained through kinetic selection only (*i.e*. in the absence of a DCL). Within the dynamic library, the recognition-enabled products *cis*-**75** and *cis*-**67** were produced much faster than the recognition-disabled counterparts produced by the reaction with nitrones **76** and **77**. Nevertheless, little change was observed in selectivity between the two recognition-mediated processes in the dynamic *vs* kinetic selection. The pyridine unit in nitrone **74** contains an additional methyl group relative to that in nitrone **65**, and, therefore, binds the carboxylic acid on **66** more strongly. This stronger association allows the recognition-mediate reaction process leading to *cis*-**75** to outperform the reaction leading to *cis*-**67**. The lack of change in the selectivity for *cis*-**75** over *cis*-**67** suggested that it is not possible to amplify the slight advantage afforded to the system with a slightly higher association constant by dynamic environment beyond the limit imposed by kinetic selection. Employing kinetic fitting of the experimental data available and through subsequent kinetic simulations, the authors explored what limits the degree of selectivity in this dynamic recognition-mediated network. The simulations showed that substantial selectivity for one product over another formed *via* two similar recognition-mediated reaction processes (proceeding *via* the AB pathway in this case) is possible only if the ratio of the association constants as well as rate constants governing these processes is higher than that for the dynamic system examined experimentally. The simulations further showed that an optimum concentration window exists for achieving selectivity. Ultimately, however, the simulations revealed that at an optimum concentration and with the right kinetic and thermodynamic parameters, a system embedded in a dynamic environment can achieve higher selectivity than an identical system governed purely by kinetic selection.

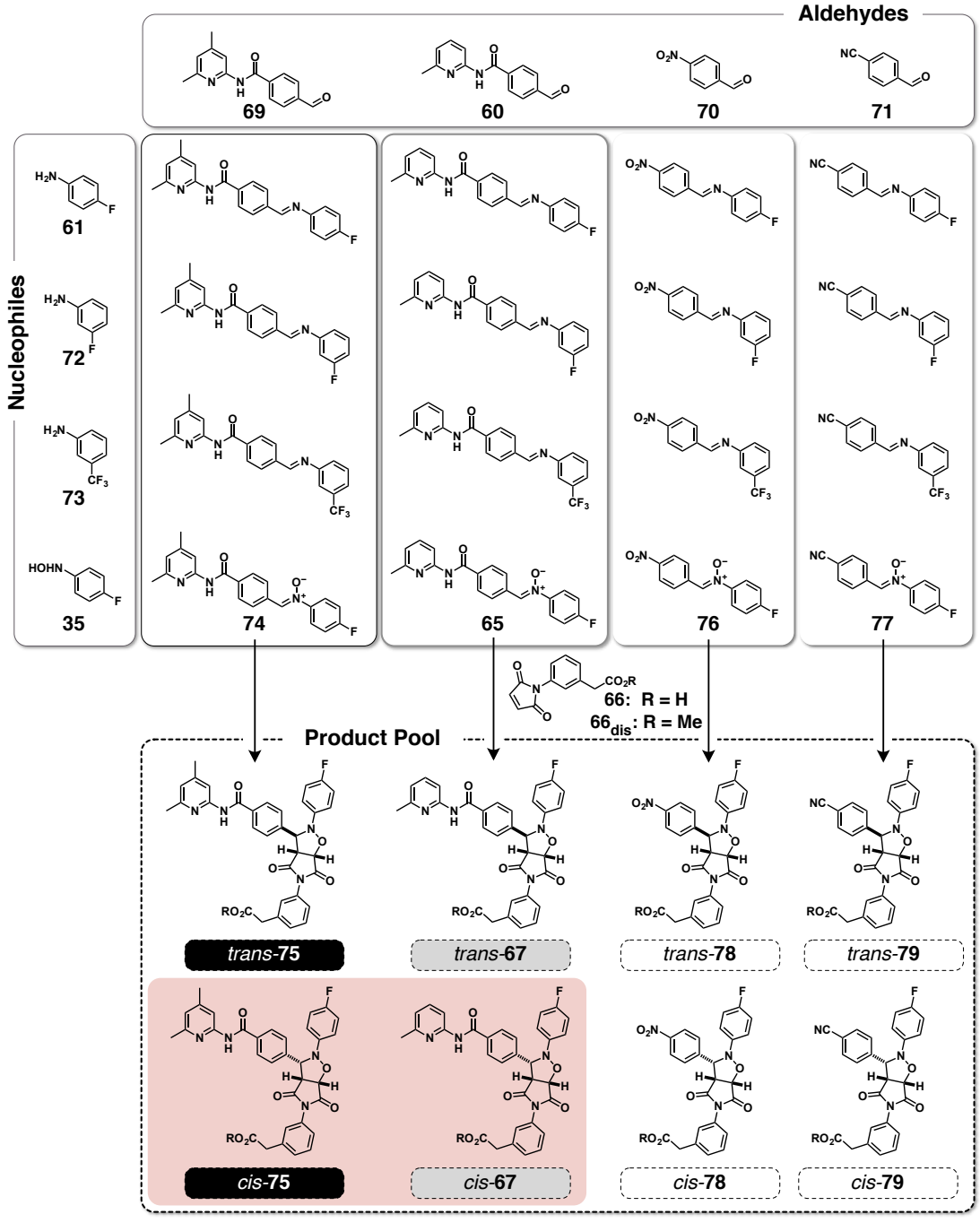


Figure 1.46 A dynamic library comprised of four aldehydes (**69**, **60**, **70** and **71**), and four nucleophiles (**61**, **72**, **73** and **35**). In addition to these building blocks, the exchange pool contains 12 imines and 4 nitrones. The nitrone components have the capacity to react with maleimide **66**, however, the reaction is recognition-mediated only for nitrone **74** and **65**. Figure adapted from Ref. 181.

Sadownik and Philp elaborated[126] on the simple four component DCL instructed with a maleimide capable of removing[180] selected library components *via* an irreversible recognition-mediated AB pathway to a library coupled to a self-replicating process. In this case, the position of the carboxylic acid on the maleimide components was altered to promote the template-mediated reaction *via* the ternary complex. Employing the *para*-substituted maleimide **M1**, previously shown[127] to form an efficient *trans* diastereoselective self-replicating system, the authors showed that instruction of the imine-nitrone dynamic library (**Figure 1.47**) resulted in rapid transformation of the recognition nitrone **65**.

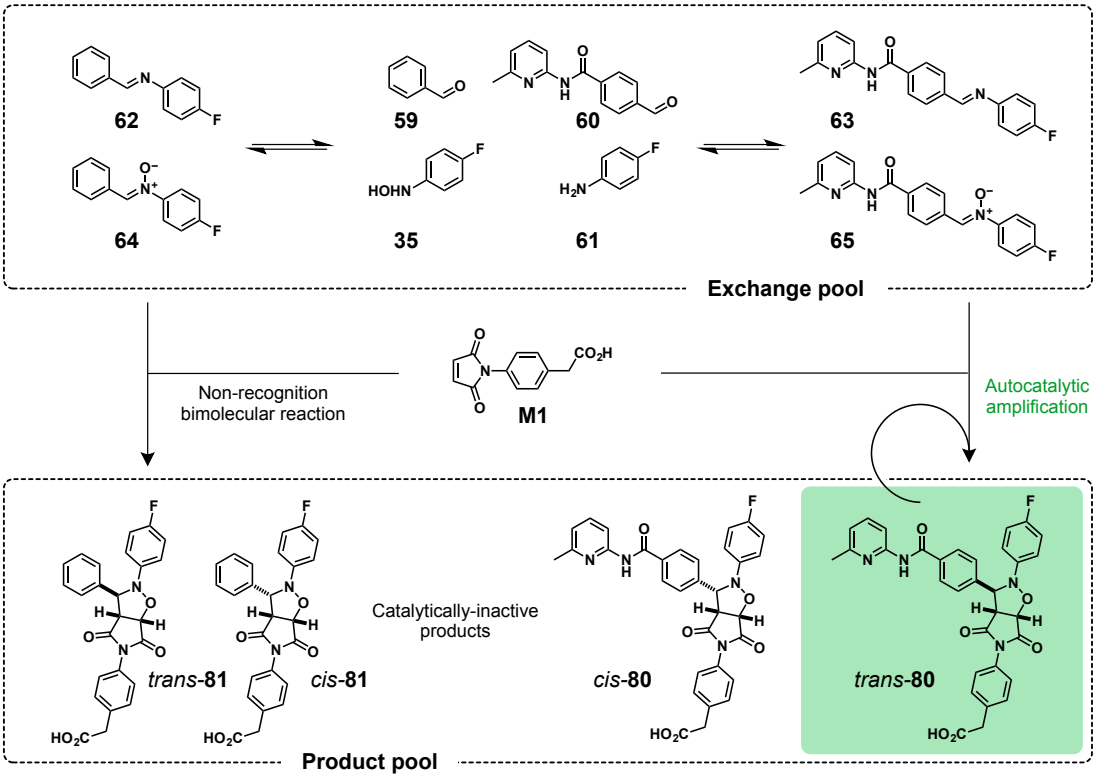


Figure 1.47 A dynamic library comprised of two interconverting imine (**62** and **63**) and two nitrone (**64** and **65**) components. Within the library, two components are equipped with a reactive nitrone site and two with an amidopyridine recognition site. However, only a single component, **65**, possess both the recognition and reactive elements required for the reaction with maleimide **M1** *via* a template-directed self-replicating pathway, forming the *trans*-**80** (green box) rapidly and with high diastereoselectivity. Nitrone **64** reacts with **M1** to form *cis*-**81** and *trans*-**81** only *via* the slow, unselective bimolecular pathway. Figure adapted from Ref. 126.

In fact, after 16 hours, more than 48% of the library was transformed to products, with *trans*-**80** (**Figure 1.47**, red box) being by far the most dominant product at 80% (formation of *cis*-**80** was not observed). Nitrone **64**, on the other hand, formed the *cis*-**81** and *trans*-**81** only very slowly and with low diastereoselectivity. The level of selectivity for *trans*-**80** was magnified further, when the library was examined in the presence of 10

mol% of preformed *trans*-**80** as instructive input. This addition of preformed template amplified the production of *trans*-**80** further, in particular by enabling the system to perform *via* the template-mediated pathway from the onset of the reaction.

1.6 Summary and outlook

Networks are the building blocks in the world around us, and the inherent interconnectedness at their core often exerts an influence over our everyday lives that is challenging to predict. The all-pervasive presence of networks in nature is reflected in the number of disciplines increasingly involved in the study of complex systems, with properties that are beyond those exhibited by their individual components. The field of chemistry also experienced a paradigm shift around twenty years ago, moving away from the study of molecular matter in isolation and starting to embrace the notion of complexity and complex networks. The interdisciplinary research directed towards the construction and analysis of complex networks, known collectively as systems chemistry,[43,63,66,182] brings together aspects of supramolecular chemistry, origins-of-life research and far-from-equilibrium systems, endeavouring to develop a better understanding of complex systems and phenomena, and the requirements leading to their emergence. Taking inspiration from the systems found in nature, in particular the complexity of living systems, systems chemistry employs the bottom-up approach to design, development and investigation of synthetic chemical systems as models for complex behaviour, with the view to examining the system-level properties that arise as a result of the interactions and reactions between the components within these networks. A phenomenon of particular interest to systems chemistry is self-replication—the complex process at the very core of living organisms.

Synthetic systems present a unique opportunity for systems chemists to study the complex phenomenon of self-replication using chemical networks constructed using molecules with well-defined structures and with catalytic and recognition properties that can be probed and characterised experimentally. The developments in the field of systems chemistry have, as illustrated by the examples presented in this introduction, produced a great variety of replicating systems based on oligonucleotides,[79,82,83,85,87] peptides[88,92,93,97,99,105] and small synthetic molecules,[109,114,115,120,127] highlighting and demonstrating[40,67,69,70,183] that the ability to replicate is not exclusive to complex biological systems, exploiting a complex enzymatic machinery. In addition to self-replication, these systems were shown to express a number of properties that emerged as a consequence of the interactions embedded in the network—error-correction, stereo-specific replication and Boolean logic operations, to name a few. Peptide-based replicating

systems in particular, have achieved[68,106,145,184,185] a notable level of sophistication, examining networks comprising more than a single replicator—a feature significantly less well-developed in replicating systems exploiting oligonucleotides and small-organic molecules.

The requirements for the operation of self-replicating systems in isolation have been established. The processes in complex networks in the real world, however, never operate in isolation—nor do complex biological systems function using fully preformed components. Instead, components of biological networks achieve their formation from mixtures of precursors. The reaction environment is, in fact, a parameter that is only beginning to be explored in systems chemistry, where majority of replicating systems are examined under the well-established, closed system conditions (*i.e.* well-stirred batch reactor model). In this respect, the DCC approach presents[134,178] an extremely useful tool for the construction of complex networks with an added component of a dynamically-exchanging pool of components—a reaction environment for the study of chemical networks that is one step closer to the dynamic, often heterogeneous environment found in nature. Nevertheless, despite the significant progress in the coupling of kinetically-driven irreversible reaction processes to dynamic covalent systems, attempts at integrating self-replication processes with the DCC approach are only just starting to appear. More scarce yet, are experimental reports directed at investigating more than a single template-mediated processes under dynamic conditions—an environment for the examination of simultaneously operating replicating systems that can facilitate our understanding of the process and the requirements that allowed a replicating species to exploit a mixture of components for its own synthesis during the processes of chemical evolution.

The aim and direction of the work described in this thesis is to exploit the expertise developed over the years in the Philp laboratory in designing self-replicating systems, and explore networks of replicator based on small organic molecules—in particular, replicators connected by a requirement for a shared building block. Such networks of interconnected replicators present a model system for studying how the reaction and recognition-mediated processes govern the preference, *i.e.* the selectivity, for one replicator over another. Building on the study of a reaction network in isolation, *i.e.* in a closed reaction environment driven by kinetic selection, the behaviour of the network can be examined within a dynamic environment, constructed using the DCC approach—allowing determination of how the dynamic selection alters the outcome of the competition between replicators.

The work will exploit various algorithms for network resolution[c]—*i.e*. the resulting distribution of replicators competing for a shared building block will be determined by the type of reaction environment and the covalent and non-covalent selection processes employed. The study of replicating networks will be aided by computational modelling, kinetic fitting and kinetic simulations—tools intended to complement the experimental analysis and allow examination of the experimental systems under a range of conditions that might be challenging experimentally. Ultimately, it is envisaged that examination of different modes of selection in a replicating system, driven by either a change in the reaction environment or the presence of an additional recognition-mediated processes, will help build a more comprehensive picture of how the network of replicators behaves, its requirements and limitations.

Some of the work described in **Section 1.5.3** has been published in:

- T. Kosikova, H. Mackenzie and D. Philp, *Chem. Eur. J.* **2016**, *22*, 1831–1839

[c]Traditionally, resolution refers to a chemical process by which a racemic mixture is separated into the constituent enantiomers. In this thesis, the term resolution will be used to denote the process by which the reaction format, the covalent and non-covalent selection processes in a network of replicators determine the final product distribution.

CHAPTER 2

GENERAL DESIGN PRINCIPLES AND OBJECTIVES

2.1 Preamble

The examination of a network of replicators constructed from small organic molecules is unlikely to identify absolutely the precise steps in the route leading to the emergence of the first self-replicating molecule capable of sustaining itself from a mixture of chemical components on the prebiotic earth. Nevertheless, studying the phenomenon of self-replication with model chemical compounds can provide information about the behaviour of a replicating system, how it correlates with its structural features, and the interplay between the various recognition and reaction processes operating within the network. In this section, the general design principles and objectives pertaining to the replicating networks examined within this thesis are introduced, with particular focus on the reaction and recognition requirements necessary for successful operation of replicating systems, and the different modes of system resolution and reaction environment available for their analysis.

2.2 The recognition event

The concepts of self-replication, autocatalysis and crosscatalysis were introduced in **Chapter 1**. The ability of a replicator to take part in template-mediated formation of itself, and in the formation of another molecule through a crosscatalytic pathway, relies on the presence of recognition elements engineered within its structure. The strength of the recognition mediating the various interactions in a system affects the capacity of a particular template molecule to initiate and participate in auto- and crosscatalytic pathways. Specifically, a single point recognition event in a replicating system drives

the association of an unreacted building block with the replicator template. Therefore, the strength of this association determines the quantity of template that needs to be formed through the slow, bimolecular pathway, before the assembly of individual components in a catalytically-active complex is possible, enabling efficient replication. The template formed through the reaction of the building blocks within the ternary complex possesses recognition motifs identical to those within the building blocks, and hence, the recognition processes exert influence also over the stability of template duplexes, affecting the level of product inhibition in each system.

As noted in the discussion of the minimal model of self-replication in **Chapter 1**, adjustments in the temperature at which a replicating system is examined can be exploited to modulate the strength of the recognition processes. For example, the interaction between recognition sites becomes stronger at lower temperatures, resulting in an increase in the proportion of all recognition-mediated complexes in solution. Amongst the recognition-mediated complexes that a decrease in temperature would affect is also the product duplex—in particular, the temperature decrease would increase the stability of this duplex, result in lower catalytic turnover. For this reason, it is imperative to consider simultaneously the benefit of increasing the proportion of catalytically-active complexes in a system and the downside of enhancing product inhibition. Adjusting the concentration of reagents can similarly alter the efficiency of the recognition event—any recognition-mediated processes will not work efficiently at concentrations below the K_d for each recognition process ($K_d = 1/K_a$).

Within the Philp laboratory, one particular recognition motif, based on the hydrogen-bonding-mediated association between an amidopyridine and a carboxylic acid has been explored extensively in the design of self-replicating systems. The strength of the association (K_a) in this recognition motif was studied using ^{1}H NMR titrations (0 °C and 10 °C, $CDCl_3$, **Figure 2.1**) by varying[186] the substituents on the amidopyridine unit and the position of the carboxylic acid on the aromatic ring: 4-bromophenylacetic acid and 3-bromobenzoic acid. **Figure 2.1** demonstrates a number of principles[187] derived from the study that can be exploited in the design of recognition-mediated replication networks and their analysis. Namely, the location of the carboxylic acid can alter the strength of the recognition quite significantly—in each case, the four examined aldehydes associated more strongly with the 3-bromobenzoic acid than with the 4-bromophenylacetic acid. The higher K_a values determined for the 3-bromobenzoic acid than for the 4-bromophenylacetic acid are in line with the pK_a values reported[188] for these carboxylic acids in the literature (pK_a benzoic acid = 4.21 and pK_a phenylacetic acid = 4.31). Similarly, the methyl groups exert a positive inductive effect on the electron density on the amidopyridine ring—the K_a increases with the number of

methyl groups present. The position of a methyl group on the pyridine ring can also alter the basicity—a methyl group in position 6- on the pyridine ring results in a slightly higher association constant with a carboxylic acid than a methyl group in 4- position. Again, this observation correlates with the basicity of methyl-substituted pyridines—the pK_a value of pyridine (pK_a = 5.25) increases in the presence of even a single methyl group: the pK_a of 2-methylpyridine is 6.06[189] and the pK_a of 4-methylpyridine is 5.99.[189] The presence of two methyl groups further increases the basicity—the pK_a of 2,4-dimethylpyridine is *ca.* 6.75.[190]

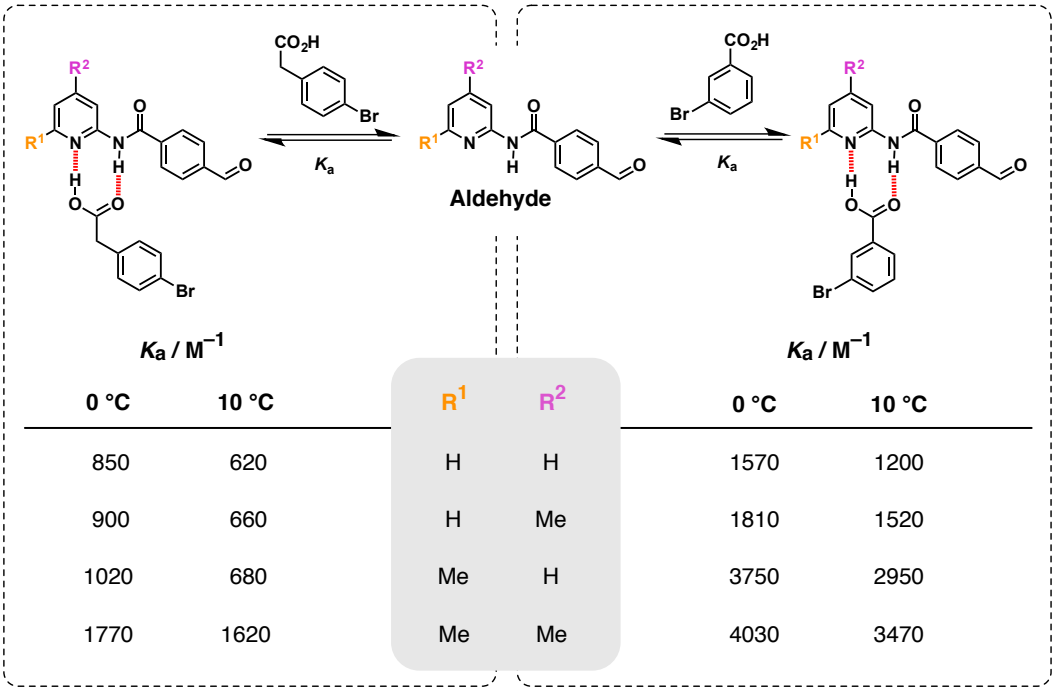


Figure 2.1 Association constants determined by ^{1}H NMR (499.9 MHz) titration for the association between various amidopyridine-based aldehydes and 3-bromobenzoic (right) and 4-bromophenylacetic acid (left), examined at 10 mM, 0 °C and 10 °C, in $CDCl_3$. Data taken from Ref. 186.

The NMR titration experiments confirmed that an increase in temperature reduces the strength of association in all of the examined amidopyridine–carboxylic acid pairs, demonstrating the possibility of altering the proportion of the complexed and free species within a system, and, thus, also the efficiency of processes relying on the formation of catalytically-active complexes. Interestingly, the variation in the number of methyl groups on the amidopyridine unit affected not only the strength of its association with the carboxylic acids but also the solubility properties of the final replicating system, incorporating these recognition units. For example, the amidopyridine unit containing no methyl groups exhibited the weakest interaction and also the worst solubility properties, whereas the opposite was true for the 4,6-amidopyridine-based replicators. The replica-

tors examined in this thesis will be based on the well-established recognition between a 6-methylamidopyridine—benzoic acid and 6-methylamidopyridine—phenylacetic acid—pairs of recognition motives that should provide a compromise between sufficient solubility and strength of association, whilst limiting the problem of product inhibition.

2.3 The chemical reaction

The recognition-mediated formation of a self-replicating template requires the joining of two individual building blocks within the ternary complex [**A**·**B**·**T**]. In an ideal situation, the product duplex formed by the reaction would be less stable than the ternary complex itself, resulting in a replicator with a high catalytic turnover. An ideal reaction suitable for the implementation in a synthetic replicating system should (i) not require inert reaction conditions or the presence of a catalyst, (ii) have a low bimolecular reaction rate constant (*i.e.* the reaction should proceed inefficiently in the absence of recognition elements), (iii) be irreversible in order to facilitate kinetic analysis and characterisation of the products formed and (iv) should be easy to follow by standard spectroscopic methods at a reasonable time-scale and temperatures. A reaction that fulfils these requirements particularly well is the 1,3-dipolar cycloaddition reaction between a nitrone and a maleimide—a reaction that has been employed successfully in the design of numerous self-replicating systems by the Philp laboratory over the last 10 years.

1,3-Dipolar cycloaddition reactions fall into a larger group of pericyclic reactions, and describe the reaction between a dipole, with a general structure that can be described[191] as a–b–c, with charges located appropriately to create a dipolar structure (**Figure 2.2a**) and a dipolarophile (*e.g.* a simple alkene, as illustrated in **Figure 2.2a**). Since their first discovery, 1,3-dipolar cycloaddition reactions have been developed into an extremely useful synthetic tool,[192,193] particularly for the synthesis of five-membered heterocyclic compounds.

The reactivity and selectivity in 1,3-dipolar cycloaddition reactions[191,194] can be explained using the frontier molecular orbital (FMO) theory. Based on the dominant set of frontier molecular orbital (FMO) interactions between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) (**Figure 2.2b**), 1,3-dipolar cycloaddition reactions can be divided[195,196] into three types: I, II and III. Using an example reaction between an alkene as a dipolarophile, and a dipole, type I is dominated by FMO interaction between the $HOMO_{dipole}$ and the $LUMO_{alkene}$. In type II reactions, the interactions between both sets of orbitals are important ($HOMO_{dipole}$ and $LUMO_{alkene}$, and $HOMO_{alkene}$ and $LUMO_{dipole}$). 1,3-Dipolar cycloaddition reac-

tions classified as type III are predominantly governed by interactions between the $HOMO_{alkene}$ and the $LUMO_{dipole}$. Generally, the reactivity[197,198] of various dipoles and dipolarophiles, can be tuned through changes in the electronic properties of the reactants, as well as through the presence of a lewis acid, which can co-ordinate either to the dipole or to the alkene, altering the energy of the FMOs.

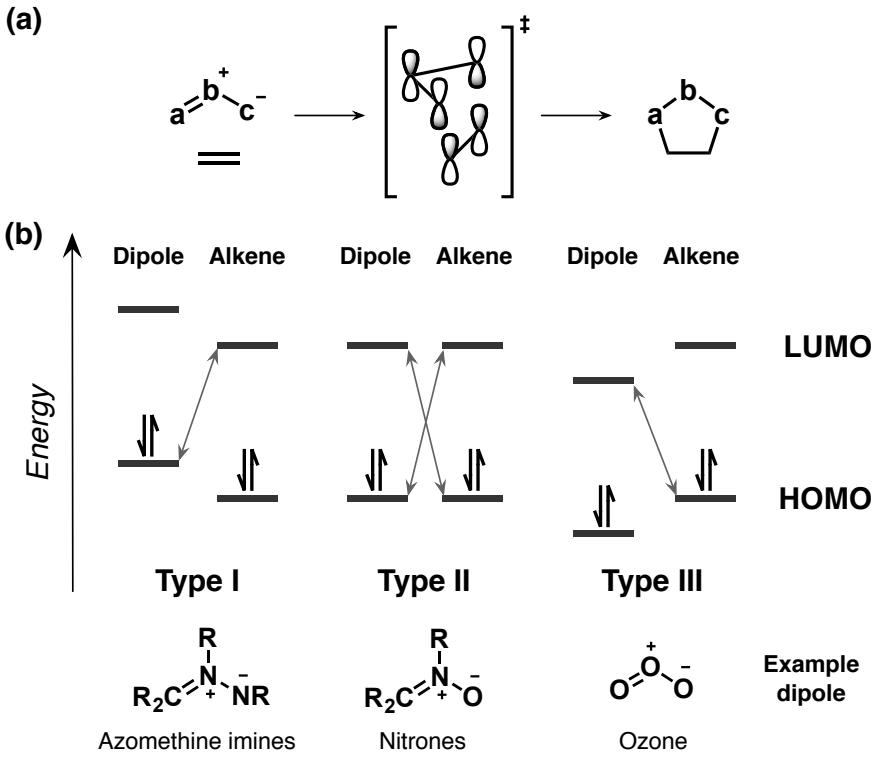
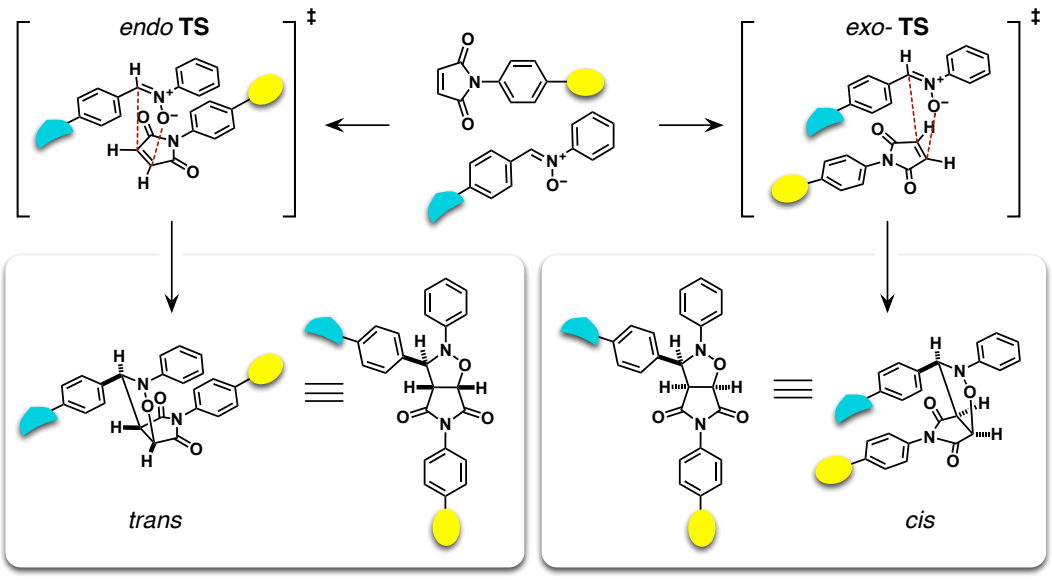


Figure 2.2 **(a)** A general representation of a 1,3-dipolar cycloaddition reaction between a dipole, defined as an a–b–c structure and an dipolarophile, giving rise to a five-membered ring. The reaction proceeds generally through a concerted transition state governed by the frontier molecular orbitals of the reactants. Figure adapted from Ref. 191. **(b)** Classification of 1,3-dipolar cycloaddition reactions based on the frontier molecular orbitals (FMOs) of the reactants (dipole and alkene) controlling the transition state. Reaction between nitrones and maleimides can be classified generally as Type II, which means that both HOMO-LUMO interactions play a role. Figure adapted from Ref. 194.

In the absence of molecular recognition, 1,3-dipolar cycloaddition reactions proceeds with a fairly low diastereoselectivity, generally affording a mixture of the *trans* and *cis* diastereoisomeric products ([*trans*]/[*cis*] ratio of *ca.* 3), formed through the *endo* and the *exo* transition state (TS), respectively (**Scheme 2.1**). Öztürk and Coşkun have investigated[197] the diastereoselectivity in product formation in the reactions between a variety of *N*-aryl maleimides and *N*-phenyl nitrones. In this study, electron-withdrawing groups on the maleimide component were found to favour the *cis* diastereoisomer (formed *via exo* TS), whereas electron-donating groups enhanced selectivity for the *trans* diastereoisomer (*endo* TS). The variation in diastereoselectivity (*i.e.* the ratio

of [*trans*]/[*cis*]) in 1,3-dipolar cycloaddition reactions as a result of changes in the electronic environment of the maleimide remained relatively low—[*trans*]/[*cis*] ratio of around 1.7 for highly electron withdrawing and 0.5 for electron donating groups—a variation mostly influenced by π–π stacking interactions between the phenyl rings of the maleimide and the nitrone components.

The inherently low diastereoselectivity makes the 1,3-dipolar cycloaddition reaction between a nitrone and a maleimide an ideal choice for the use in self-replicating systems. In a reaction lacking any recognition elements, the ratio of [*trans*]/[*cis*] diastereoisomers will remain low. However, if the reactants in the system are designed to contain recognition sites that enable these components to take part in stabilising interactions, which allow the formation of catalytically-active complexes, the selectivity for a particular diastereoisomer can increase significantly. For example, **Scheme 2.1** illustrates the example of a replicating system with recognition sites located on the maleimide *N*-aryl ring and on the nitrone aryl ring (in position away from the nitrogen atom of the nitrone).



Scheme 2.1 A 1,3-dipolar cycloaddition reaction between a nitrone (orange) and a maleimide (green) has two possible diastereoisomeric products: *trans* and *cis*. In this case, the *trans* diastereoisomer (formed *via* the *endo* TS) only possesses the recognition sites in a suitable conformation to allow the assembly of building blocks in a catalytically-active ternary complex. The recognition components in the *cis* diastereoisomer, formed *via* the *exo* transition state, are positioned close together, rendering this product catalytically-inert.

In this case, the maleimide and the nitrone can react *via endo* and *exo* transition states. However, only the *trans* diastereoisomer possesses the two recognition sites in an 'open' conformation (*i.e.* the recognition sites pointing away from each other), necessary for its assembly with the unreacted building blocks in a catalytically-active ternary complex that facilitates their reaction through the *endo* TS. The recognition

components in the *cis* diastereoisomer, formed *via* the *exo* transition state, are positioned in close proximity to each other, rendering this product catalytically-inert. The lack of diastereoselectivity in the 1,3-dipolar cycloaddition reaction between a nitrone and maleimide in the absence of any recognition elements, can, therefore, be used as a tool for quantifying the effect that self-replication has on the diastereoselectivity within a particular system—a useful measure of the efficiency of structural information transfer in a system that exploits template-mediated processes for its formation.

2.4 Algorithm for network resolution

A network formed by two replicators that share a common building block can be directed to form one of the replicators by addition of preformed template (*i.e.* with an increased selectivity relative to the condition lacking the added template). The network resolution depends on the selection processes operating in a particular network—*i.e.* the algorithms (**Figure 2.3**). Traditionally, in a system where all replicators are formed through an identical irreversible step, the transformation of reaction building blocks into products is determined by a single recognition-mediated covalent capture process (**Figure 2.3**, algorithm A)—the variation in the strength of the recognition-mediated processes and the rate of the covalent bond forming step will affect directly the ability of the replicators to compete for a shared building block.

An alternative algorithm for network resolution can be envisaged in a situation where, akin to the examples of dynamic covalent systems coupled to replication presented in **Chapter 1**, the building blocks required for the production of the replicators are distributed within a dynamic covalent library, from which the self-replicating species have to accomplish their formation. In addition to the recognition-mediated covalent capture process, the selection in a system of replicators embedded in a dynamic pool of components depends also on the covalent selection processes (**Figure 2.3**, algorithms A and B)—the building blocks required for the formation of replicators have to first be formed from their constituent dynamic components. Such algorithm introduces another level of control that contributes towards the overall outcome in a competition network—the so-called dynamic selection.

The algorithm of network resolution can be altered further by introducing an additional, alternative, non-covalent recognition-mediated process (**Figure 2.3**, algorithm C) that operates in the system in parallel. In this case, the resolution is driven simultaneously by the recognition-mediated covalent capture and the non-covalent selection processes (**Figure 2.3**, algorithms A and C) and the interplay between them. A specific example where the algorithm for network resolution is directed by two such algorithms can be

envisaged in an environment leading to the formation of a mechanically-interlocked architecture, such as a [2]rotaxane, capable of self-replication. Of particular interest in such a network, driven by the addition of a macrocycle and the presence of a suitable macrocycle binding site within the molecular framework, is the examination of the interplay between the selection driven by the recognition-mediated covalent capture processes required for replication and those permitting the formation of a mechanically-interlocked rotaxane, and its effect on the overall product distribution.

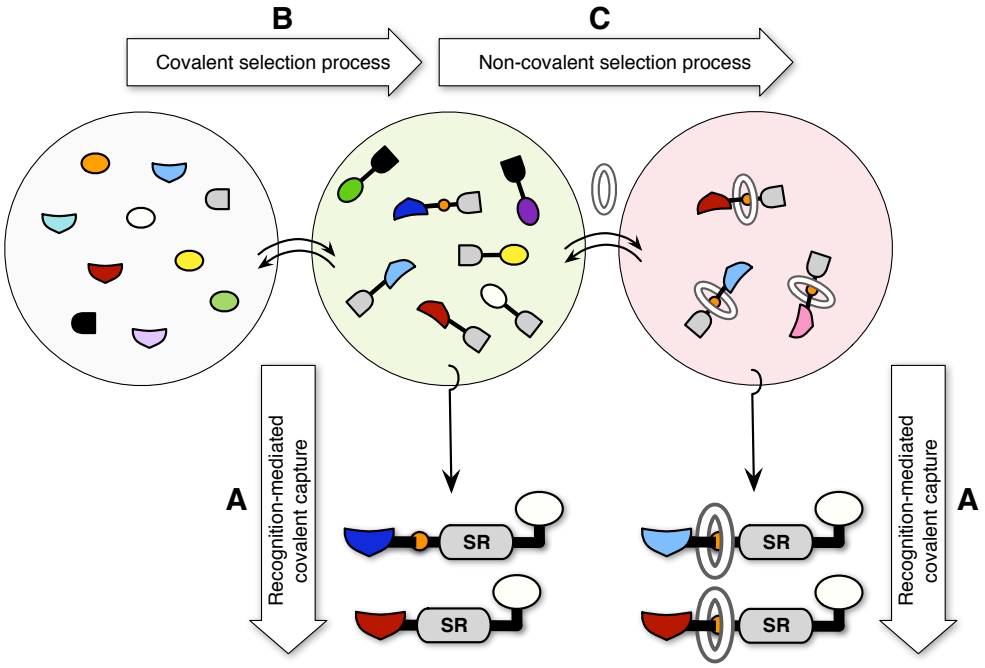


Figure 2.3 The selectivity in a network of self-replicators (SR) competing for a shared building block depends on the algorithm for network resolution employed. For example, a network can be constructed from fully preformed building blocks where selection is driven by a recognition-mediated covalent capture process (algorithm A). Alternatively, the reaction components can be distributed amongst the components of a dynamic covalent library (DCL), from which the replicators have accomplish their formation (algorithm B)—an environment where both non-covalent and covalent selection operates (algorithms A and B). In addition to the variation of the components, selection for a particular replicating process over another, and thus also the overall resolution of a network, can be examined in the presence of another non-covalent selection process (algorithm C)—conditions where network resolution is driven by the recognition-mediated non-covalent capture as well as the addition non-covalent selection process (algorithms A and C).

These three algorithms (**Figure 2.3**, algorithms A, B and C) for network resolution presented thus far—(i) a scenario involving a single recognition-mediated covalent capture process (algorithm A), (ii) a network driven by covalent and non-covalent selection processes simultaneously (algorithms A and B), and (iii) a system where selection is driven by multiple recognition-mediated selection processes operating in parallel (algorithms A and C)—are all examples of a reaction environment where replication takes place in a homogeneous, well-stirred media within a closed reaction environment

(*i.e.* well-stirred batch reactor format, **Figure 2.4a**). In addition to the non-covalent and covalent selection processes driving the network resolution, the selectivity for one replicator over another depends on the reaction environment. In the three examples of closed reaction environment (**Figure 2.3**, A to C), the selectivity in a network where two replicators share a building block is limited by the inability of the replication processes to perform at their optimum efficiency throughout the duration of the reaction. In essence, the finite concentration of the reagents available for the formation of the replicators becomes self-limiting—as the replicators exhaust the building blocks available for their formation, the efficiency of replication processes decreases, reducing the degree of selectivity between two competing recognition-mediated processes at the same time.

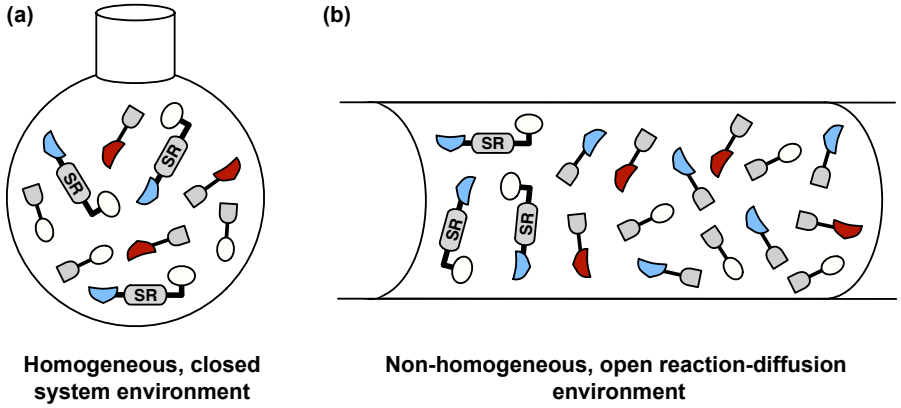


Figure 2.4 Graphical representation of replication in **(a)** a well-stirred batch reactor environment and **(b)** under reaction-diffusion conditions, where a propagating reaction-diffusion front of autocatalytic template can be initiated by localised addition of preformed template.

The inherent barrier to selectivity for one replicator over another imposed by the closed nature of the reaction environment can be overcome by exploiting conditions where the self-replication never becomes self-limiting—conditions far-from-equilibrium. A suitable vehicle for the examination of replication systems under such 'open' conditions can be found in the format of reaction-diffusion systems (**Figure 2.4b**). In a reaction-diffusion system, addition of a small amount of preformed solution of a particular replicator, at a specific location in an expanse of reactants, generates a concentration gradient, allowing the autocatalytic template to initiate the formation of a propagating reaction-diffusion front as it diffuses away from the region of high template concentration. The reaction environment provided by the reaction-diffusion conditions should permit the replication processes to perform at their optimum efficiency throughout the reaction—as the replicator diffuses through space, it should encounter mostly unreacted building blocks. The same principle can be applied to the examination of networks of competing replicators under reaction-diffusion conditions—the system

resolution can move beyond the limits imposed by closed reaction environment, allowing the more efficient replicator to extinguish the weaker replicator over time. The reaction-diffusion format opens up the possibility of examining replicating systems under non-homogeneous out-of-equilibrium conditions, more similar to those present in nature.

Bringing together the presented types of algorithms for network resolution and the various reaction environments available for the examination of replicating systems, it is clear that the selectivity for one replicator over another can be affected by altering the non-covalent and covalent selection processes within a network, as well as the nature of the reaction format. The algorithm for network resolution based on a single non-covalent selection process in a network assembled from preformed components constitutes an example of a closed reaction environment that is employed the most commonly for the analysis of replicating systems—it is also a scenario that permits the baseline for selectivity between competing recognition-mediated, irreversible processes to be established in a homogeneous, well-stirred environment. Nevertheless, only comparison of the effects that these alternative algorithms for resolution exert in a replicating system will allow a comprehensive set of the parameters that govern and limit the selectivity in networks of competing replication processes to be formulated.

2.5 Summary

The main objective of the work presented in this thesis is to explore various algorithms and reaction formats for resolving a network of competing replicators. The individual components used to construct a replicating systems will be designed to possess the capacity to interact and react together *via* recognition-mediated processes, and the system-level behaviour emerging as a result of the embedded interactions will be examined. Networks of replicators will be studied with the view to exploring the effect of structural variations in the components comprising the system on the capacity of the replicators to perform a certain function, such as the ability self-replicate and to interact with other components within the network (for example through crosscatalytic pathways). The efficiency of replication in these interacting networks will be analysed through kinetic experiments and characterisation of the individual recognition and reaction processes. The work in the thesis will first examine the behaviour of a network of two competing replicators experimentally as well as through kinetic simulations, under closed reaction environment conditions, driven by recognition-mediated covalent capture—*i.e*. kinetic selection (**Chapter 3**, algorithm A). In the next step, the same network will be examined within the environment of a dynamically exchanging library

of components (**Chapter 4**, algorithms A and B) and under reaction-diffusion conditions (**Chapter 5**, algorithm A). Finally, one of the replicators comprising this network will be modified to integrate another non-covalent selection process, in addition to the recognition-mediated covalent capture processes driving self-replication—in this case a recognition process required for the assembly of a mechanically-interlocked architecture within the same system (**Chapter 6**, algorithm A and C).

CHAPTER 3

A KINETIC STUDY OF AN INTERCONNECTED NETWORK BASED ON TWO SYNTHETIC REPLICATORS

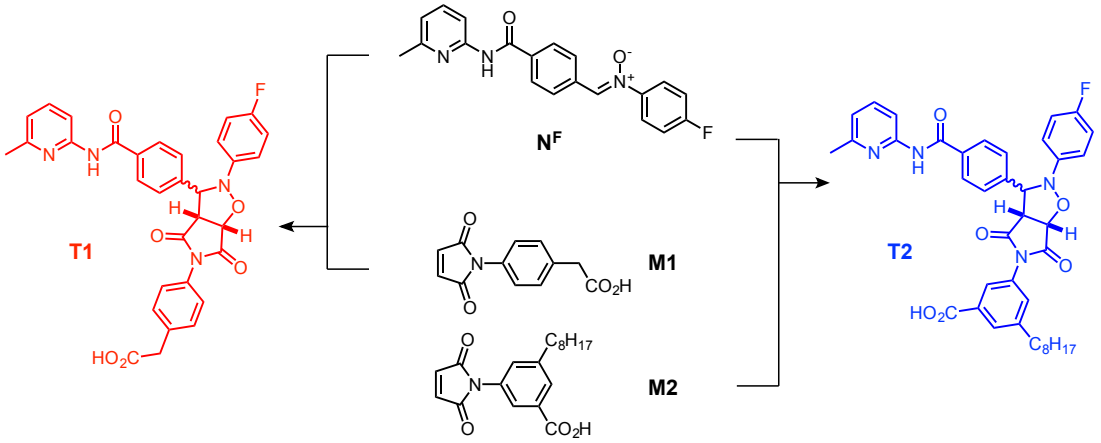
3.1 Preamble

Numerous self-replicating systems based on small molecules have been reported in the literature to date. However, majority of the work focuses on the kinetic analysis of individual replicators, and detailed investigations of interconnected networks of replicators are lacking. The study of systems incorporating more than one replicator opens up the possibility of multiple auto- and crosscatalytic cycles, giving rise to catalytic and inhibitory relationships between replicators, which in turn can produce complex feedback loops that are more similar to those observed in biological networks. Biological systems exhibit the capacity for selective replication, *i.e*. a phenomenon where a single replicating molecule is amplified selectively, with high fidelity. In synthetic small-molecule based networks of replicators, selective replication remains a challenge.

3.2 Aims and objectives

The aim of this chapter is to undertake a comprehensive kinetic analysis of an interconnected system of two structurally-similar self-replicators, and to investigate how the selectivity for one replicator over another changes in the presence of instructing preformed template. The two replicators are connected by a shared nitrone building block, $\mathbf{N^F}$, which can react with a maleimide **M1** and **M2** through 1,3-dipolar cycload-

dition reactions[a] to form two replicators: **T1** and **T2** (**Scheme 3.1**). Previous work in the Philp laboratory has demonstrated[126,199] that both of these replicators are capable of establishing efficient autocatalytic cycles. Nevertheless, these two replicators, **T1** and **T2**, have thus far never been examined under competition conditions—a situation where both maleimides are present in the reaction mixture simultaneously, competing for a limited amount of a shared nitrone starting material. Similarly, the ability of **T1** and **T2** replicators to template the synthesis of each other *via* the crosscatalytic pathways has not been investigated and will be probed through template instructed kinetic experiments in this chapter. Finally, the comprehensive kinetic analysis will be supplemented by kinetic fitting and simulations with the view to probing the resolution of a network of **T1** and **T2** replicators, thereby establishing the limits on selectivity in the competing replication processes in a closed reaction format, driven by kinetic selection.



Scheme 3.1 Schematic representation of **T1**–**T2** replicating network formed by 1,3-dipolar cycloaddition between **M1** and **M2** maleimides with nitrone $\mathbf{N^F}$. Throughout this chapter, **T1** cycloadduct and the recognition-processes pertaining to it will be represented in red, while blue colour will be used to represent **T2**.

3.2.1 Recognition and reaction processes in the system

Replicators **T1** and **T2** are equipped with two recognition elements: an amidopyridine moiety and a carboxylic acid group. The main distinguishing feature of these two replicators is the position of the carboxylic acid group, located on each maleimide. In maleimide **M1**, and thus also replicator **T1**, the carboxylic acid recognition site is a

[a]The 1,3-dipolar cycloaddition reaction between a nitrone and a maleimide can give two diastereoisomers: *trans* and *cis*. Only the *trans* diastereoisomer is capable of taking part in template-directed replication processes in the network of **T1** and **T2** replicators, and, thus, the notation of the cycloadduct products capable of replication will generally omit the *trans* notation throughout this thesis (for example, *trans*-**T1** will be generally referred to as **T1**), unless emphasis on the identity of the diastereoisomer is necessary.

phenylacetic acid present in position *para* relative to the maleimide ring (**Figure 3.1**). Maleimide **M2**, in comparison, is equipped with a benzoic acid functionality, in position *meta* with respect to the maleimide ring—a feature that is conserved in replicator **T2** (**Figure 3.1**). The 6-methyl amidopyridine recognition site, originating from the shared nitrone building block, is identical in both replicators and was selected out of the various amidopyridine recognition units possible for its favourable solubility properties and the large difference between the K_a values determined for its association with a phenylacetic acid and a benzoic acid based compounds. Therefore, the difference in the recognition processes within each replicator system, as well as the variation in the length of the formed template as a result of the position of the carboxylic acid, was envisaged to produce a network of two replicators, referred to as the **T1**–**T2** network, with interesting crosscatalytic behaviour.

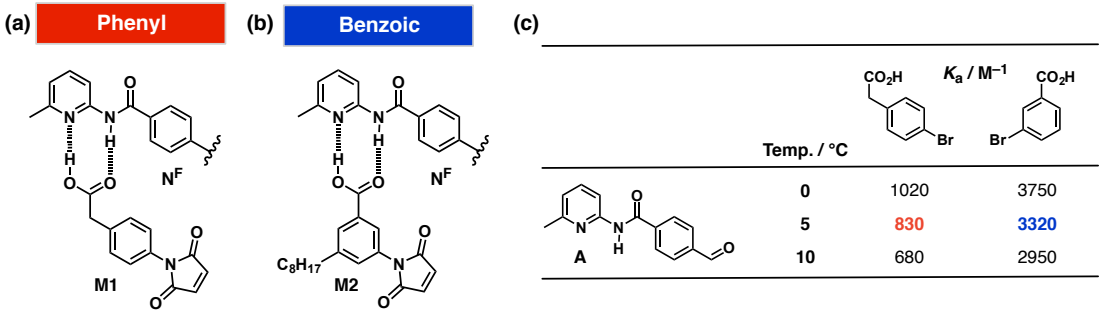


Figure 3.1 The hydrogen-bonding-mediated recognition between a 6-methyl amidopyridine moiety on nitrone $\mathbf{N^F}$ and carboxylic acid recognition site present in **(a)** phenylacetic acid **M1** maleimide and **(b)** benzoic acid maleimide **M2**. **(c)** Association constants (K_a / M^{-1}) determined for the interaction between aldehyde **A** (analogue of $\mathbf{N^F}$) and 4-bromophenylacetic acid (analogue of **M1**) and 4-bromobenzoic acid (analogue of **M2**) at 0 °C and 10 °C. K_a values for temperature 5 °C were determined through van't Hoff plot and are highlighted in red (**T1**) and blue (**T2**).

The formation of the two replicators relies on the 1,3-dipolar cycloaddition reaction between a nitrone and a maleimide. The minimal model of self-replication was introduced and described in detail in **Chapter 1**. The maleimide components bear a COOH recognition group which allows these components to associate with the amidopyridine moiety on the $\mathbf{N^F}$. The single point recognition event can mediate the formation of binary complexes [**M1**·$\mathbf{N^F}$] and [**M2**·$\mathbf{N^F}$] which can produce a 'closed' template, *cis*-**T1** and *cis*-**T2**, *via* the recognition-mediated binary complex reaction pathway. These closed *cis* products possess no free recognition sites and are thus catalytically-inactive. Alternatively, the maleimides can react with the nitrone through an uncatalysed and template-independent bimolecular pathway, to give both the *cis* and *trans* diastereoisomeric products of each replicator. Only the *trans* cycloadducts possess the two recognition sites in an 'open' conformation, required for the successful

assembly of the maleimide and nitrone components in to the catalytically-active ternary complex [**T**·**M**·**N**$^{\mathbf{F}}$] (where **M** denotes a maleimide).

The slight structural difference in the recognition element in **T1** and **T2**, crucial for the ability of these systems to take part in recognition-mediated reactions, is expected to manifest in their ability to template their own synthesis. Bringing the individual components, required for the formation of the two replicators together permits the formation of a system that can form both products simultaneously, and is responsive to addition of external stimulus (addition of preformed template), opening up the possibility of creating an instructable replicating network (**Figure 3.2**) that can be directed to make a particular replicator with increased selectivity, *i.e.* preferentially compared to the condition with no instruction. Nevertheless, the ability of each replicator to compete for the limited shared building block, and, therefore, the selectivity for one replicator over another will depend on the efficiency of the autocatalytic cycles present in the system as well as the activity and efficiency of the two crosscatalytic pathways available to the system.

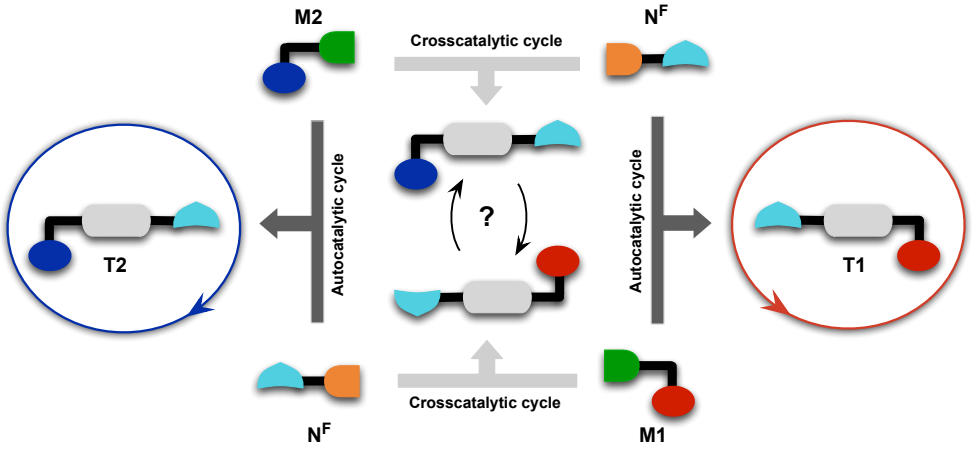


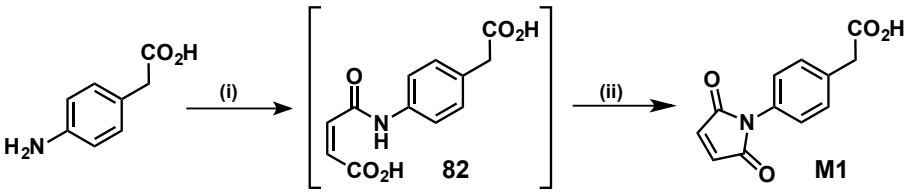
Figure 3.2 Cartoon model of the **T1**–**T2** replicating network, composed of three building blocks: maleimides **M1** and **M2**, and nitrone **N**$^{\mathbf{F}}$. The *trans*-**T1** and *trans*-**T2** replicators formed by the reaction of the nitrone with each maleimide.

In order to understand fully the catalytic efficiencies of the replicators, formation of **T1** and **T2** replicators will be examined in isolation first, *i.e.* reaction of nitrone with one maleimide at a time. In order to establish the relative contribution of each bimolecular pathway, the kinetic analysis of each replicator will be undertaken using a recognition-disabled model maleimide. Subsequently, the kinetic behaviour of each replicator will be examined in isolation, under various conditions: in the absence of template and in the presence of preformed autocatalytic and crosscatalytic template, allowing the efficiencies of the catalytic pathways available to **T1** and **T2** replicators to be determined. Following the full characterisation of the individual catalytic pathways,

these two replicators will be examined in a competition environment, first in the absence of template and later using preformed template, added at the beginning of each reaction to instruct the network. Kinetic analysis, permitted by the presence of an aryl-fluorine tag in the nitrone $\mathbf{N^F}$, and, thus, also in each cycloadduct formed by the reaction with this nitrone, will be performed using $^{19}F\{^1H\}$ NMR spectroscopy as the key analytical method.

3.3 Syntheses of network components

The two structurally-similar replicators, **T1** and **T2**, are formed upon reaction of a fluorine-tagged nitrone $\mathbf{N^F}$ with **M1** and **M2** maleimide, respectively. The two maleimides, **M1** and **M2**, required for the synthesis of theses replicators, were prepared from commercially available 4-aminophenylacetic acid and 3-nitrobenzoic acid, respectively. Maleimie **M1** was furnished by a two-step, one-pot reaction with maleic anhydride and the subsequent immediate cyclisation of the key intermediate **82** at 120 °C, in 30% yield (**Scheme 3.2**).



Scheme 3.2 Conditions and reagents employed in the synthesis of maleimide **M1**. (i) Maleic anhydride, AcOH, RT, 4 h, quantitative conversion. (ii) AcOH, 120 °C, 30%.

Synthesis of the recognition-disabled maleimides, $\mathbf{M1^C}$ and $\mathbf{M2^C}$ (**Figure 3.3**) was not necessary as there was a sufficient stock of these compounds available from the previous members of the Philp laboratory.

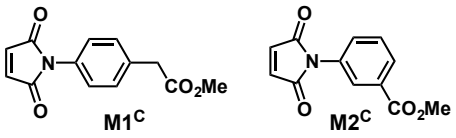
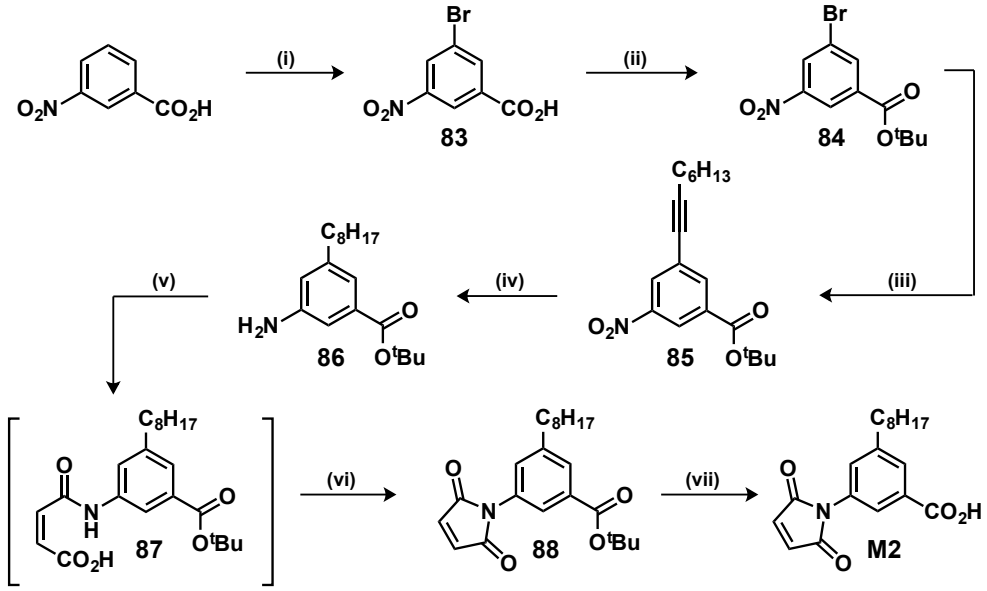


Figure 3.3 Control, recognition-disabled maleimides $\mathbf{M1^C}$ and $\mathbf{M2^C}$.

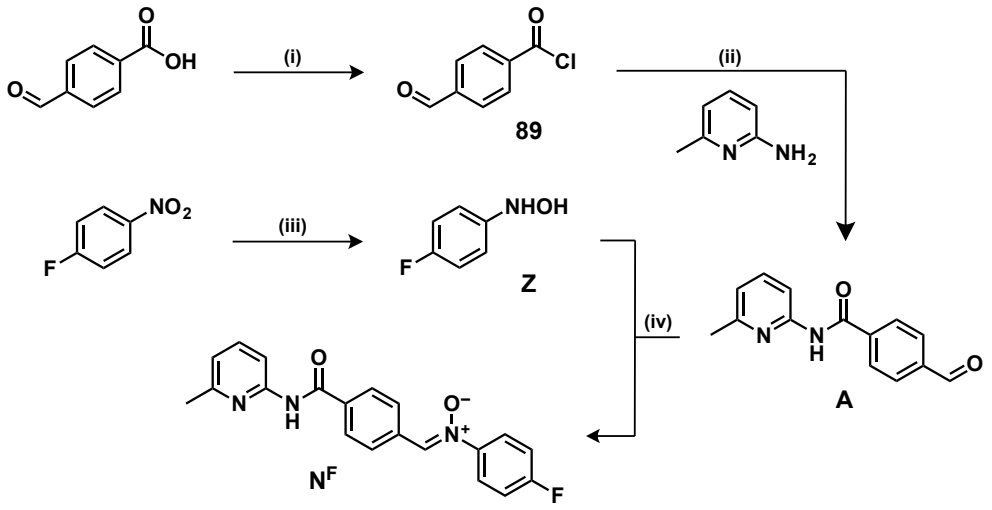
Synthesis of maleimide **M2** (**Scheme 3.3**) commenced with a selective monobromination of 3-nitrobenzoic acid, followed by protection of the brominated carboxylic acid **83** as a *tert*-butyl ester, **84**. In the next step, a Sonogashira coupling was employed in order to functionalise the aromatic ring with a alkynyl chain, which improves the solubility, yielding product **85**. Using an H-cube continuous flow hydrogenation reactor

and a palladium catalyst, the nitro group and the triple bond in **85** were reduced simultaneously in a single reaction step (over three consecutive runs), furnishing **86** in 90% yield. Reaction of **86** with maleic anhydride produced the intermediate compound **87** which was used immediately without purification. Cyclisation of **87** in the presence of Ac_2O and NaOAc at 110 °C afforded *t*-Bu protected **M2**, **88**. In the last step of the synthetic route, deprotection of the *t*-Bu ester **88** yielded the desired, recognition-enabled maleimide **M2** (49%), in 25% overall yield, starting from 3-nitrobenzoic acid.



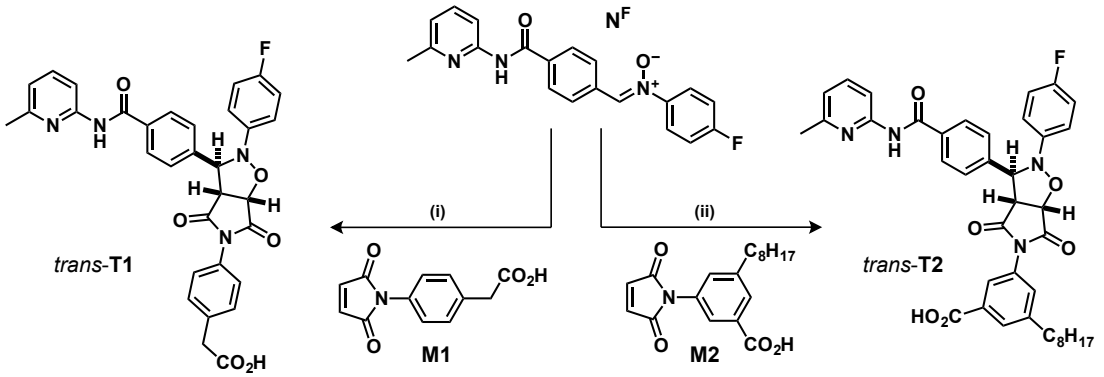
Scheme 3.3 Conditions and reagents employed in the synthesis of maleimide **M2**. (i) 1.1 eq. NBS, 60 °C, H_2SO_4, 96%. (ii) di-*tert*-butyldicarbonate, DMAP, THF, RT, 79%. (iii) CuI, PPh_3, $PdCl_2(PPh_3)_2$, Et_3N, octyne, 80 °C, 16 h, 86%. (iv) H-cube, Pd/C (10% wt), MeOH/THF (0.08 M), 50 °C, 50 bar, 3 runs, 90%. (v) Maleic anhydride, THF, RT, quantitative conversion. (vi) Ac_2O, NaOAc, 110 °C, 86%. (vii) TFA, CH_2Cl_2, RT, 49%.

The fluorine-tagged nitrone $\mathbf{N^F}$ was prepared by reaction of aldehyde **A** and hydroxylamine **Z**. Synthesis of these two building blocks began with three commercially available components: 4-formylbenzoic acid, 6-methyl-aminopyridine and 1-fluoro-4-nitrobenzene (**Scheme 3.4**). In the first step of the synthesis, 4-formylbenzoic acid was converted to its corresponding acyl chloride **89** using $SOCl_2$ in PhMe and directly coupled to 6-methylpyridin-2-amine, to afford aldehyde **A** in 53% yield. 1-fluoro-4-nitrobenzene was reduced to *N*-(4-fluorophenyl)hydroxylamine **Z** using rhodium on carbon and hydrazine monohydrate. Finally, condensation of the hydroxylamine **Z** with aldehyde **A** in EtOH afforded the desired nitrone $\mathbf{N^F}$ in an excellent 80% yield.



Scheme 3.4 Reagents and conditions employed in the synthesis of Nitrone $\mathbf{N^F}$. (i) $SOCl_2$, PhMe, 110 °C, 24 h, quantitative conversion. (ii) Et_3N, CH_2Cl_2, 0 °C to RT, 53%. (iii) THF, Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, RT, 1 h, 87%. (iv) EtOH, dark, 16 h, 80%.

With the fluorine-tagged nitrone and both maleimides readily available, the two catalytically-active *trans* templates, *trans*-**T1** and *trans*-**T2** (**Scheme 3.5**), were synthesised on a preparative scale from their constituent components, for use in the instructed kinetic experiments. Extended reaction times were employed in order to ensure complete conversion of the maleimide and nitrone building blocks to products, which precipitated out of solution over time, allowing the templates to be isolated by vacuum filtration. Synthesis of template *trans*-**T2** was slightly more challenging as a result of the presence of the *cis* diastereoisomer in the reaction mixture. Nevertheless, following recrystallisation of the crude sample of **T2**, both templates were obtained in sufficient quantity for characterisation and for use in template-instructed kinetic experiments.



Scheme 3.5 Reagents and conditions employed in the synthesis of fluorinated cycloadducts *trans*-**T1** and *trans*-**T2**. (i) $CDCl_3$, 4 °C, 13 days, 51%. (ii) $CDCl_3$, 4 °C, 13 days, 33%.

3.4 Kinetic analyses of replicators in isolation

The designed network of two replicators encompasses numerous recognition-mediated and reaction processes, forming an interconnected system of three building blocks, the products of their reaction and numerous recognition-mediated complexes. Multiple reaction pathways are available to each replicating system, and, therefore, a full kinetic analyses of the individual replicators in isolation was performed prior to examining the network under competition conditions—environment where both maleimides and nitrone are present simultaneously.

3.4.1 Recognition-disabled kinetic experiments

As the first step in the comprehensive kinetic analyses of the two replicators, the reaction of the nitrone was examined with each recognition-disabled control maleimide, $\mathbf{M1^C}$ and $\mathbf{M2^C}$—a situation where the formation of the catalytically-active ternary complexes ([$\mathbf{M1}\cdot\mathbf{N^F}\cdot\mathbf{T1}$] and [$\mathbf{M2}\cdot\mathbf{N^F}\cdot\mathbf{T2}$]) is not possible (**Figure 3.4**).

In a typical kinetic experiment, an equimolar solution of a maleimide and nitrone $\mathbf{N^F}$ (5 mM) in $CDCl_3$ was allowed to react at 5 °C. The reaction was monitored by 470.3 MHz $^{19}F\{^1H\}$ NMR spectroscopy at 30 minute intervals, over 16 hours. The concentrations of the reaction components were determined at each time point relative to 1-bromo-2-fluoro-4-nitrobenzene as an internal standard. The data obtained was used to construct concentration *vs* time profiles for the recognition-disabled kinetic experiments. The error in concentrations determined in the NMR kinetic experiments reported in this chapter was estimated at $\pm$ 0.02 mM (for details, see **Chapter 9**).

The reaction time course[b] observed for the formation of $\mathbf{T1^C}$ and $\mathbf{T2^C}$ from their constituent components (**Figure 3.4a** and **c**) shows that both 1,3-dipolar cycloaddition reactions proceed in the absence of recognition-processes in the system with low conversion and little diastereoselectivity. Specifically, reaction of $\mathbf{M1^C}$ and $\mathbf{N^F}$ forms both diastereoisomeric products very slowly, where the recognition-disabled *trans*-$\mathbf{T1^C}$ reached concentration of 0.14 mM and *cis*-$\mathbf{T1^C}$ was formed at concentration of 0.04 mM after 16 hours (**Figure 3.4**).

The recognition-disabled reaction between $\mathbf{M2^C}$ and $\mathbf{N^F}$ exhibited a reaction profile similar to the reaction with $\mathbf{M1^C}$, however, the transformation of the reaction components into the products proceeded slightly more efficiently, resulting in 0.25 mM concentration of *trans*-$\mathbf{T2^C}$ and 0.10 mM concentration of *cis*-$\mathbf{T2^C}$ after 16 hours (**Figure 3.4b** and **d**).

[b]The reaction time course shown in **Figure 3.4**, as well as a large number of the reaction profiles reported in this chapter also include the outcome of kinetic fitting of the experimental data (shown as solid lines)—these results will be discussed in more detail later in the chapter.

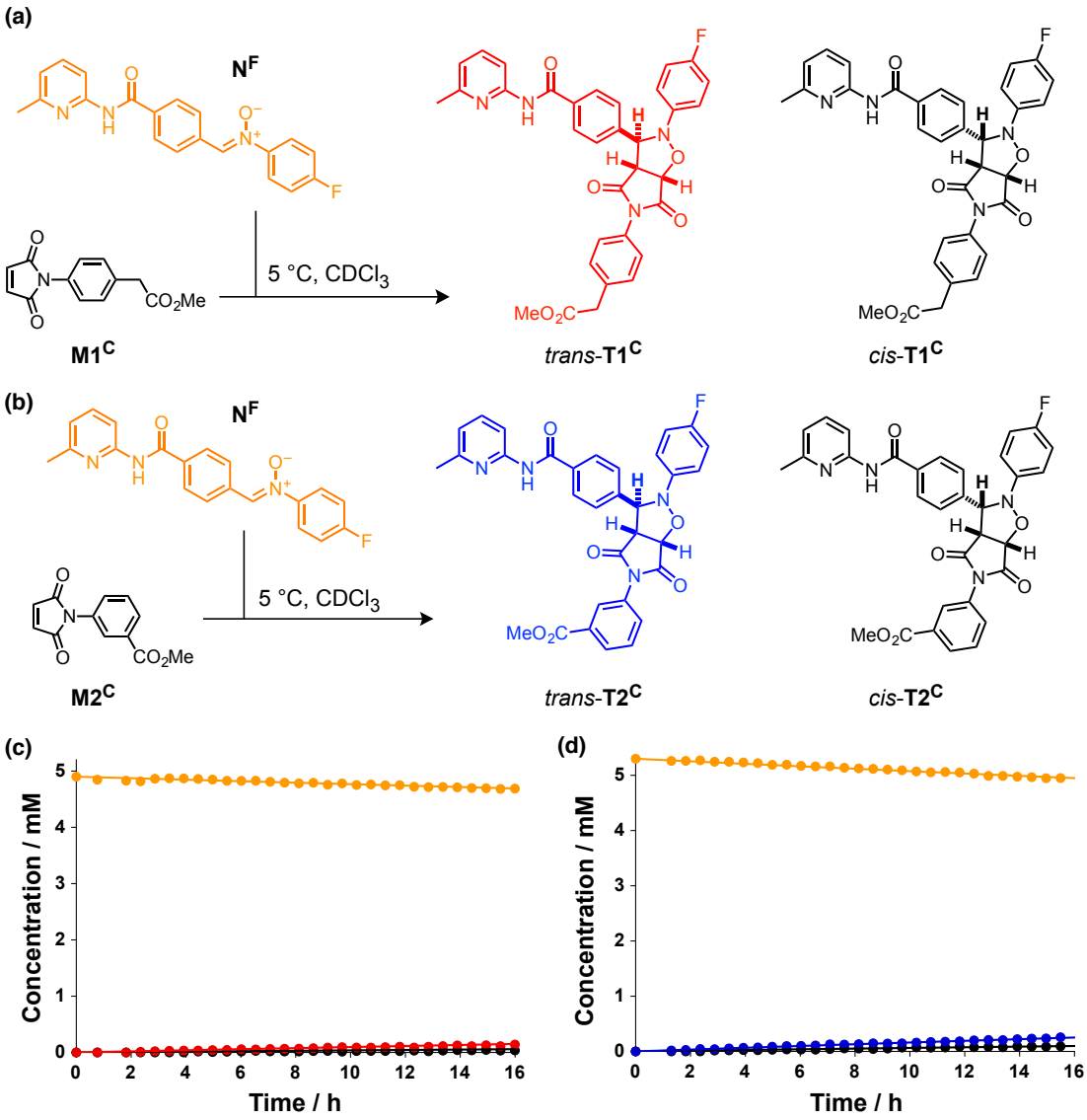


Figure 3.4 Recognition-disabled kinetic experiments examining the contribution of the bimolecular pathway to the formation of **T1** and **T2** replicators. Nitrone $\mathbf{N^F}$ reacts with **(a)** $\mathbf{M1^C}$ and with **(b)** $\mathbf{M2^C}$ to form recognition-disabled **(c)** *trans*-$\mathbf{T1^C}$ and *cis*-$\mathbf{T1^C}$ and **(d)** *trans*-$\mathbf{T2^C}$ and *cis*-$\mathbf{T2^C}$. Deconvoluted experimental data, as determined by $^{19}F\{^1\}$ NMR spectroscopy: concentrations were determined relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard, is shown in circles, while solid lines represent the result of kinetic fitting. Nitrone is shown in orange, *trans*-$\mathbf{T1^C}$ in red, *trans*-$\mathbf{T2^C}$ in blue and both *cis* cycloadducts are shown in black. Reaction conditions: 5 °C, $CDCl_3$, 16 hours.

In both recognition-disabled experiments, the combined conversion to *trans* and *cis* diastereoisomers did not exceed 7%, with a ratio of [*trans*]/[*cis*] diastereoisomeric products below 3 in each case, thereby confirming that in the absence of recognition, the 1,3-dipolar cycloaddition reactions are slow and unselective.

3.4.2 Recognition-enabled kinetic experiments

3.4.2.1 Uninstructed kinetics

In a self-replicating system, the bimolecular pathway leads to slow formation of both diastereoisomeric products. Once a sufficient quantity of the open *trans* template has been formed, the constituent components can assemble into the catalytically-active ternary complex, establishing an autocatalytic cycle. In order to confirm that *trans*-**T1** and *trans*-**T2** are capable of exploiting recognition-mediated reaction processes for their formation, the reaction between each recognition-enabled maleimide, **M1** and **M2**, with nitrone $\mathbf{N^F}$ was examined.

An equimolar solution of **M1** and $\mathbf{N^F}$ ([**M1**] = [$\mathbf{N^F}$] = 5 mM) was prepared ad the time course of the reaction was evaluated by 470.3 MHz $^{19}F\{^1H\}$ NMR spectroscopy. Formation of **T1** proceeded with high diastereoselectivity, forming the *trans* diastereoisomer exclusively. The reaction exhibited a sigmoidal reaction profile, characteristic of a self-replicating system, forming *trans*-**T1** at concentration of 0.53 mM after four hours and 3.34 mM after 16 hours (**Figure 3.5a** and **c**). Recognition-enabled formation of **T2** from its constituent components (**Figure 3.5b** and **d**) exhibited a similar sigmoidal reaction profile, although the reaction proceeded with slightly lower diastereoselectivity and efficiency. After four hours, *trans*-**T2** reached concentration of 0.30 mM and *cis*-**T2** 0.13 mM, affording the two diastereoisomeric products, [*trans*]/[*cis*], in a 2.3 ratio. After 16 hours, the concentration of these products increased to 2.10 mM and 0.40 mM for *trans* and *cis*, respectively. At this later reaction time, the *trans* to *cis* ratio increased to 5.2, highlighting the more efficient template-mediated formation of the *trans* diastereoisomer of **T2**, relative to catalytically-inert *cis* diastereoisomer, in the recognition-enabled kinetic experiment.

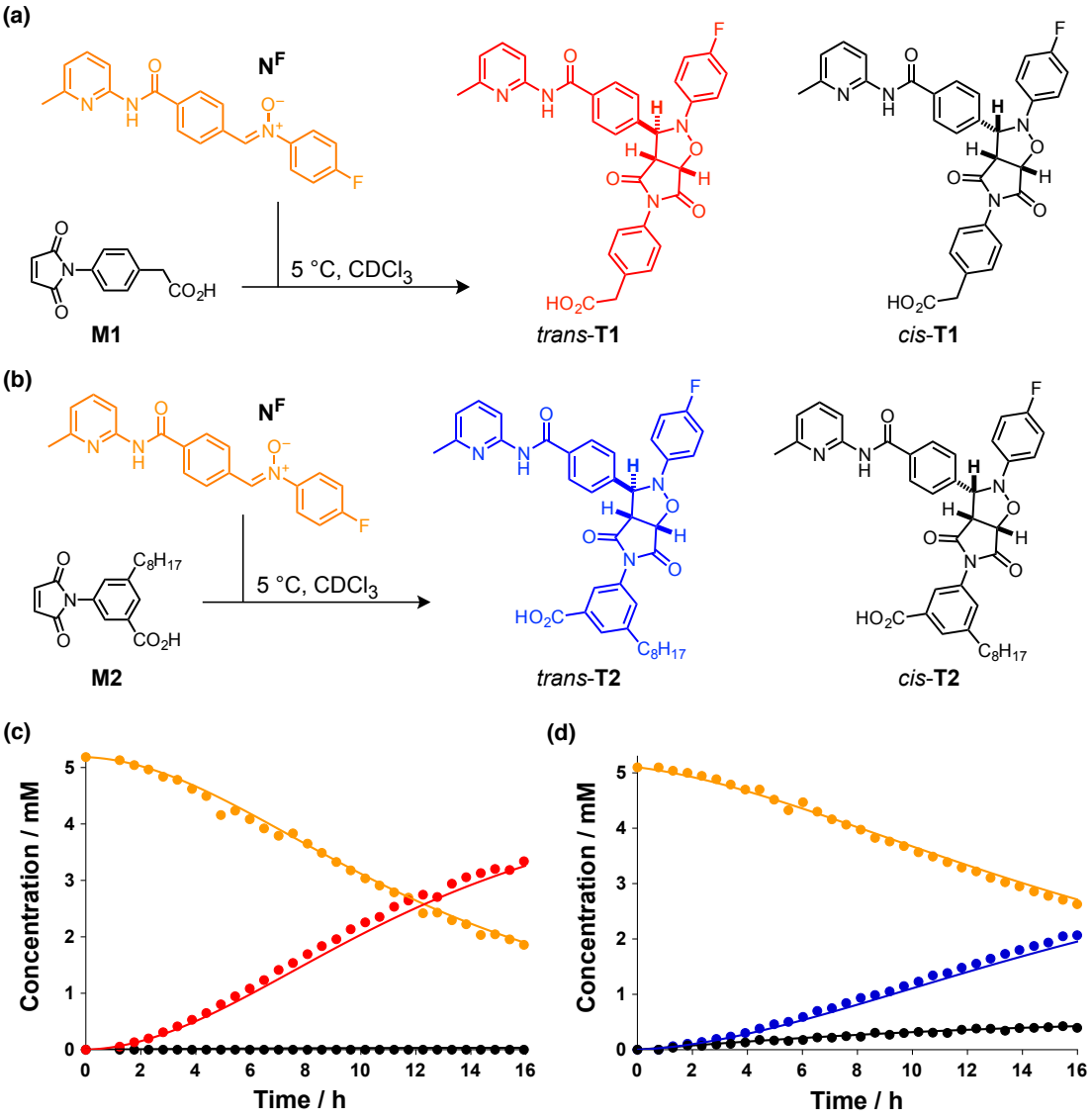


Figure 3.5 Kinetic experiments examining the formation of **T1** and **T2** replicators from their recognition-enabled building blocks. Nitrone $\mathbf{N^F}$ reacts with **(a) M1** and with **(b) M2** to form recognition-enabled **(c)** *trans*-**T1** (no *cis* was detected) and **(d)** *trans*-**T2** and *cis*-**T2**. Deconvoluted experimental data, as determined by $^{19}F\{^1H\}$ NMR spectroscopy: concentrations were determined relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard, is shown in circles, while solid lines represent the result of kinetic fitting. Nitrone is shown in orange, *trans*-**T1** in red, *trans*-**T2** in blue and both *cis* products are shown in black. Reaction conditions: 5 °C, $CDCl_3$, 16 hours.

3.4.2.2 The autocatalytic cycles

Presence of the recognition elements had a dramatic effect on the conversion and diastereoselectivity in both **T1** and **T2** replicators in the uninstructed kinetic experiments. In the next step, formation of each replicator was examined in the presence of preformed autocatalytic template (*e.g.* **T1** added to components needed for the formation of **T1**) in order to unambiguously establish their capacity to self-replicate. These instructing experiments were performed by monitoring the formation of each replicator, from the respective maleimide and nitrone components (5 mM, $CDCl_3$), in the presence of 20 mol% of preformed template (*trans*-**T1** or *trans*-**T2**, 1 mM). Progress of each reaction was monitored by 470.3 MHz $^{19}F\{^1H\}$ spectroscopy at 30 minute intervals over 16 hours. Deconvolution of the obtained kinetic data (for details, see **Chapter 9** and for an example script, see **Appendix A1**) allowed construction of a concentration *vs* time profile for **T1** (**Figure 3.6a**) and **T2** (**Figure 3.6b**) replicator.

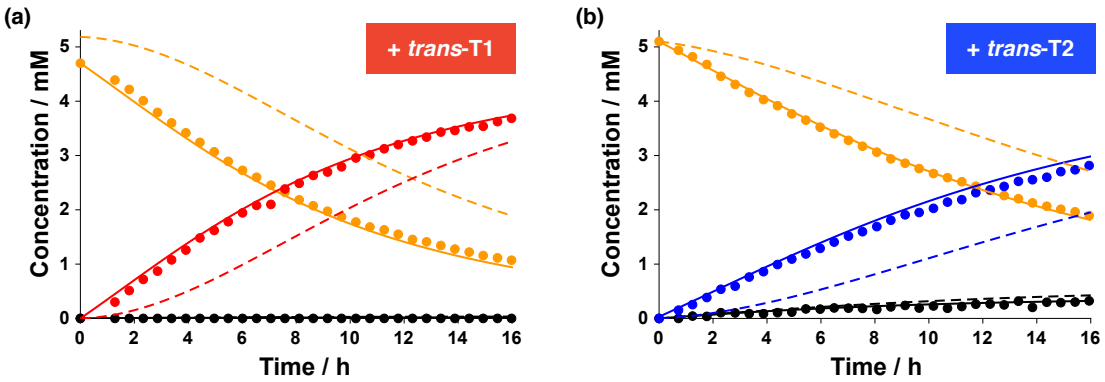


Figure 3.6 Kinetic experiments examining the formation of **T1** and **T2** replicators from their recognition-enabled building blocks in the presence of preformed autocatalytic template. Nitrone $\mathbf{N^F}$ reacts with **(a) M1** in the presence of 20 mol% of *trans*-**T1** and with **(b) M2** in the presence of 20 mol% of *trans*-**T2** to form recognition-enabled **(c)** *trans*-**T1** (no *cis* was detected) and **(d)** *trans*-**T2** and *cis*-**T2**. Deconvoluted experimental data is shown in circles, while solid lines represent the result of kinetic fitting. The results of fitting obtained in the uninstructed kinetic experiments (dashed lines) are provided for comparison. Nitrone is shown in orange, *trans*-**T1** in red, *trans*-**T2** in blue and both *cis* products are in black. Reaction conditions: 5 °C, $CDCl_3$, 16 hours, monitored by $^{19}F\{^1H\}$ NMR spectroscopy, concentrations were determined relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard. Data is corrected for the template added.

Examination of the concentration *vs* time profiles (**Figure 3.6**) constructed from these kinetic experiments examining **T1** and **T2** formation in the presence of their respective autocatalytic templates revealed the disappearance of the initial lag period in each time course, confirming the ability of each *trans* replicator to template its own synthesis, *i.e.* to self-replicate. In contrast to these uninstructed experiments, the kinetics undertaken in the presence of preformed template resulted in a noticeably higher conversion after both four and 16 hours. For example, the concentration of *trans*-**T1**

at these time points reached 1.26 mM and 3.69 mM, respectively, with no *cis* product detected. The formation of *trans*-**T2** replicator proceed slightly less efficiently in the presence of the preformed template, reaching a concentration of 0.86 mM and 2.82 mM after four and 16 hours. The catalytically-inactive diastereoisomer of this replicator, *cis*-**T2** was formed at concentration of 0.09 mM and 0.32 mM at the four and 16 hour time point (resulting in a ratio of [*trans*]/[*cis*] = 10.1 and 8.1, respectively).

3.4.2.3 The crosscatalytic cycles

The kinetic experiments described thus far showed clearly that both replicators are capable of self-replication, exhibiting a significant enhancement in both reactivity and diastereoselectivity in the recognition-enabled kinetic experiments—confirming the successful design of the two replicators. In order to determine whether **T1** and **T2** possess the capacity to crosscatalyse the formation of each other, formation of these two species was examined in the presence of 20 mol% of the corresponding preformed crosscatalytic template (*e.g.* **T1** was added to the reaction components required for the formation of **T2**).

The two crosscatalytic kinetic experiments were performed by monitoring the formation of each replicator, from the respective maleimide and nitrone components (5 mM, $CDCl_3$), in the presence of 20 mol% of preformed crosscatalytic template, (*trans*-**T2** or *trans*-**T1** (1 mM). Progress of each reaction was monitored by 470.3 MHz $^{19}F\{^1H\}$ NMR spectroscopy at 30 minute intervals over 16 hours. Kinetic analyses of the **T1** crosscatalytic pathway showed that the reaction time course exhibited a sigmoidal reaction profile with a clear lag period. Replicator *trans*-**T1** reached concentration of 0.43 mMafter four hours—a value noticeably more similar to that determined in the uninstructed as opposed to the **T1**-instructed kinetic experiment. After 16 hours, the concentration of **T1** increased to 3.15 mM (**Figure 3.7**), a value slightly lower than that observed in the uninstructed kinetic experiment. Put together, this kinetic experiment revealed that *trans*-**T2** template is not a suitable catalyst for the formation of *trans*-**T1**.

Examination of the reaction time course for the formation of *trans*-**T2** in the presence of preformed *trans*-**T1**, on the other hand, revealed a decrease in the lag period, similar to that observed in the kinetic experiment instructed with the autocatalytic template *trans*-**T2**. After four hours, diastereoisomers *trans*-**T2** and *cis*-**T2** were formed at concentrations of 0.87 mM and 0.08 mM, respectively. These concentrations increased further after 16 hours, reaching 2.52 mM and 0.22 mM, with a [*trans*]/[*cis*] ratio of 10.3 and 11.4 (**Figure 3.7**). The increased diastereoselectivity in the **T2** forming crosscatalytic experiment relative to that using preformed **T2** template suggests that the *trans*-**T1** template, which incorporates the amidopyridine and carboxylic acid recognition sites

at a slightly larger distance from each other, can promote the formation of the *trans* diastereoisomer slightly more efficiently that *trans*-**T2** template.

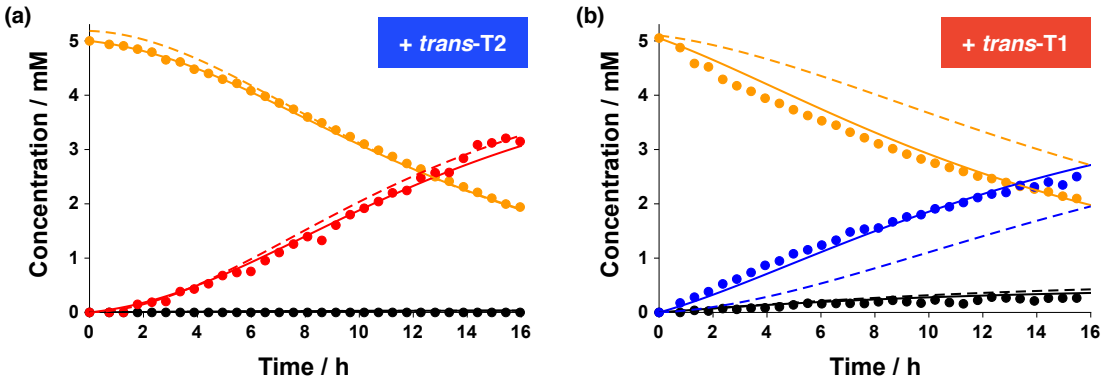


Figure 3.7 Kinetic experiments examining the formation of **T1** and **T2** replicators from their recognition-enabled building blocks in the presence of preformed crosscatalytic template. Nitrone $\mathbf{N^F}$ reacts with **(a) M1** in the presence of 20 mol% of *trans*-**T2** and with **(b) M2** in the presence of 20 mol% of *trans*-**T1** to form recognition-enabled **(c)** *trans*-**T1** (no *cis* was detected) and **(d)** *trans*-**T2** and *cis*-**T2**. Deconvoluted experimental data is shown in circles, while solid lines represent the result of kinetic fitting. The results of fitting obtained for the uninstructed kinetic experiments (dashed lines) are provided for comparison. Nitrone is shown in orange, *trans*-**T1** in red, *trans*-**T2** in blue and both *cis* products are shown in black. Reaction conditions: 5 °C, $CDCl_3$, 16 hours, monitored by $^{19}F\{^1H\}$ NMR spectroscopy, concentrations were determined relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard.

3.4.2.4 Percentage enhancements in template instructed experiments

In order to quantify the changes in replication efficiency in the template instructed experiments relative to the uninstructed counterparts, % enhancements (**Equation 9.1**) were determined for each product using the conversion results reached after 16 hours (**Figure 3.8a**) in each kinetic experiment.

$$\%Enhancement = 100 \cdot \frac{[\mathrm{T}]_{instructed} - [\mathrm{T}]_{uninstructed}}{[\mathrm{T}]_{uninstructed}} \qquad (\mathbf{3.1})$$

In the presence of 20 mol% of preformed *trans*-**T1**, formation of this product, determined for t = 16 hours, proceed with 10 ± 1% enhancement (**Figure 3.8a**). In contrast, the enhancement for *trans*-**T2** formed in the **T2** instructed experiment showed a significantly higher % enhancement[c] of 36 ± 2% (**Figure 3.8**). The higher enhancement observed for the **T2** replicator at the examined time point arises from the fact that unlike *trans*-**T1**, formation of this product is less limited by the decreasing concentration

[c]The error values for the calculated % enhancements were determined using standard methods for calculation of normally-distributed errors. These calculations employed the estimated value of concentration error of ± 0.02 mM determined for all of the kinetic experiments described in this chapter (for details, see **Chapter 9** and **Appendix A6**).

(availability) of the reagents required for its formation—*i.e.* **T2** replicator is further away from its plateau than **T1**. This conclusion is corroborated by the examination of the % enhancements observed for these two species after four hours, which clearly show dramatically higher values for both instructed experiments (**Figure 3.8b**). Specifically, 138 ± 7% and 185 ± 15% enhancements have been determined for *trans*-**T1** and *trans*-**T2** replicators, respectively.

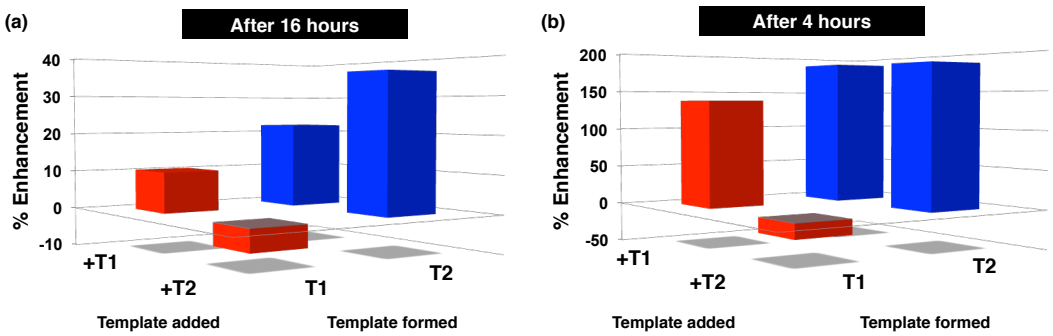


Figure 3.8 Percentage enhancement for the formation of *trans*-**T1** and *trans*-**T2**, calculated for the two autocatalytic and two crosscatalytic pathways, relative to the uninstructed experiments after **(a)** 16 hours and **(b)** four hours. All experiments were carried out in $CDCl_3$, at 5 °C. Errors for the calculated % enhancements are reported in the main text.

Similar analysis was undertaken for the crosscatalytic experiments (**Figure 3.8**). In the presence of 20 mol% of preformed *trans*-**T2**, formation of *trans*-**T1** showed negative enhancement of −6 ± 1% (t = 16 hours). The decrease in replication efficiency of *trans*-**T1** suggests that *trans*-**T2**, equipped with complementary recognition sites, sequesters **M1** and $\mathbf{N^F}$ within unproductive complexes and **T1** within the [**T1**·**T2**] heteroduplex, thereby inhibiting the formation of *trans*-**T1** relative to its formation in the absence of any template. The decrease in replication efficiency was even more marked at t = 4, −19 ± 5%, most likely as a result of the stronger affinity of the *trans*-**T2** template for $\mathbf{N^F}$ than *trans*-**T1** template for $\mathbf{N^F}$. In contrast, the enhancement determined for *trans*-**T2** formed in the **T1** instructed experiment exhibited a % enhancement of 22 ± 1% (t = 16 hours) (**Figure 3.8**). This positive enhancement, while lower than that observed in the kinetic experiment instructed with *trans*-**T2**, shows that *trans*-**T1** can efficiently template the formation of *trans*-**T2**. The enhancement is even higher, 187 ± 16%, at t = 4 hours, suggesting that the *trans*-**T1** template can instruct the formation of *trans*-**T2** more efficiently than *trans*-**T2** itself—at this reaction time, the **T1** template is the dominant product in solution, as opposed to the template **T2**, which is capable of forming stronger complexes. Overall, kinetic analyses of the **T1** and **T2** replicators showed that only three catalytic channels, out of the four possible pathways available to the system, operate efficiently.

3.4.2.5 Kinetic fitting: catalytic efficiency and template duplex stability

Kinetic analyses of **T1** and **T2** replicators revealed a clear difference in the catalytic abilities of these two templates. Exploiting kinetic fitting, the differences in the kinetic behaviour of these replicators can be rationalised in terms of the rate and association constants specific to the reaction and recognition processes governing formation of each replicator through the various catalytic pathways. Using SimFit simulation and fitting software,[83] each set of kinetic data from the recognition-disabled kinetic experiments was fitted to an appropriate kinetic model encompassing the interactions and reactions leading to the formation of **T1** and **T2** through the bimolecular pathway. The interactions in each kinetic model were transformed by the SimFit program into series of rate equations, ultimately affording the bimolecular reaction rate constant (k_{bi}) for the formation of both *trans* and *cis* diastereoisomeric products for each replicator in the recognition-disabled experiments. The bimolecular rate constants extracted through the fitting procedure, together with the association constant values for the individual binding event (K_a^{Ind}) between the aldehyde and each maleimide (**Figure 3.1c**), were employed in the fitting of the recognition-enabled data. Specifically, the kinetic data from the uninstructed experiment and that instructed with the autocatalytic template were fitted simultaneously to a single kinetic model for each replicator. The fitting procedure allowed determination of two additional kinetic parameters for each replicator, namely the recognition-enabled rate constant (k_{uni}) and the *trans* template duplex association constant (K_a^{Duplex}).

The bimolecular and unimolecular rate constants were used to calculate the kinetic effective molarity[200–202] ($EM_{kinetic}$) and the free energy of connection (ΔG^S) for each replicator. $EM_{kinetic}$ (**Equation 3.2**) provides a measure of (i) the enhancement in the template-mediated pathway relative to the bimolecular pathway and (ii) the concentration at which the reaction would have to be carried out at, in order for the bimolecular pathway to perform at the same efficiency as the recognition-mediated pathway. The latter point is the reason why rate acceleration in the template-instructed pathway still takes place even if $EM_{kinetic} < 1$, under conditions where the reaction concentration is $<$ $EM_{kinetic}$. The parameters EM_{thermo} (**Equation 3.3**) and ΔG^S (**Equation 3.4**) represent a measure of the [Template·Template] duplex stability (K_a^{Duplex}), relative to the stability of the corresponding ternary complex [$\mathbf{T}\cdot\mathbf{M}\cdot\mathbf{N^F}$], mediated by the association constant for the association between each pair of recognition elements (K_a^{Ind}).

$$EM_{kinetic} = \frac{k_{uni}}{k_{bi}} \qquad (\mathbf{3.2})$$

$$\mathrm{EM}_{\mathrm{thermo}} = \frac{K_{\mathrm{a}}^{\mathrm{Duplex}}}{(K_{\mathrm{a}}^{\mathrm{Ind}})^2} \tag{3.3}$$

$$\Delta G^{\mathrm{S}} = \mathrm{RT}\ln\frac{K_{\mathrm{a}}^{\mathrm{Duplex}}}{(K_{\mathrm{a}}^{\mathrm{Ind}})^2} \tag{3.4}$$

Further information, describing in more detail the process of kinetic fitting, calculation of kinetic parameters and an example kinetic fitting script, is available in the experimental section of **Chapter 9** and in the **Appendix A2**. Kinetic parameters extracted for the **T1** and for **T2** replicator using this fitting procedure are shown in **Table 3.1**.

Table 3.1 Overview of kinetic parameters, obtained using SimFit, for the self-replicating reaction between nitrone $\mathbf{N^F}$ and maleimide **M1** ([**M1**] = [$\mathbf{N^F}$] = 5 mM, [*trans*-**T1**] = 1 mM) or **M2** ([**M2**] = [$\mathbf{N^F}$] = 5 mM, [*trans*-**T2**] = 1 mM) at 5 °C, in $CDCl_3$).

	T1		**T2**	
	cis	*trans*	*cis*	*trans*
k_{bi} / 10^{-4} M^{-1} s^{-1}	0.431	1.12	0.704	1.79
k_{uni} / 10^{-4} s^{-1}	–	71.6	0.0231	17.0
$\mathrm{EM}_{\mathrm{kinetic}}$ / M	–	64.0	0.0328	9.47
$K_{\mathrm{a}}^{\mathrm{Duplex}}$ / 10^6 M^{-1}	–	17.9	–	68.6
$\mathrm{EM}_{\mathrm{thermo}}$ / M	–	26.0	–	6.22
ΔG^{S} / kJ mol^{-1}	–	7.53	–	4.23
$K_{\mathrm{a}}^{\mathrm{Ind}}$ / M^{-1}	830		3320	

The kinetic parameters determined through the fitting procedure provide a means of quantifying the differences in replication efficiency determined for **T1** and **T2** replicators throughout the various kinetic experiments. Both diastereoisomeric products of replicator **T2**, *trans* and *cis*, are formed with a higher bimolecular rate constant than the corresponding products in the **T1** system. The differences in the bimolecular reaction rates are likely to affect the time needed for each system to produce a sufficient amount of template needed for the assembly of the building blocks in catalytically-active complexes, dependent on the $K_{\mathrm{d}}^{\mathrm{Ind}}$ for the specific recognition processes governing the assembly ($K_{\mathrm{d}}^{\mathrm{Ind}} = 1/K_{\mathrm{a}}^{\mathrm{Ind}}$). In addition to the higher bimolecular rate constant, replicator **T2** exhibits also a higher $K_{\mathrm{a}}^{\mathrm{Ind}}$ for the recognition between the carboxylic acid and the amidopyridine unit, and, therefore, might be expected to be capable of establishing an efficient self-replicating cycle at a lower concentration of template than the *trans*-**T1** replicator. However, it is possible that the capacity of the two replicators might vary slightly from the behaviour predicted by the recognition-disabled kinetic experiments, as the maleimide reaction components employed within these two sets of kinetic experiments are structurally similar, but not identical.

T2 replicator is less diastereoselective for the formation of the *trans* product than **T1** replicator, suggesting that the reactive ternary complex [$\mathbf{N^F}$·**M1**·**T1**] assembles the maleimide and nitrone in a conformation that provides a better alignment of their reactive sites than the alignment of reactive sites available to the [$\mathbf{N^F}$·**M2**·**T2**] complex.

The initial advantage provided to **T2** replicator as a result of the higher bimolecular rate (**T2** replicator should reach the concentration required for initiating the autocatalytic cycle earlier in the reaction time course than **T1** replicator) is countered by the slower recognition-mediated rate constant for the formation of **T2**. Comparison of the **T1** and **T2** kinetic effective molarities and kinetic profiles confirm that **T1** is a more efficient replicator than **T2** template in isolation. The *trans* homoduplex association constants for the *trans*-**T1** and *trans*-**T2** replicators were determined as $17.9 \times 10^6\ \text{M}^{-1}$ and $68.6 \times 10^6\ \text{M}^{-1}$, respectively. The higher stability of [**T2**·**T2**] suggests that formation of **T2** replicator is hindered by product inhibition more significantly than replicator **T1**. Interestingly, the higher duplex stability is enough to slow down the efficiency of replication of **T2**, even though **T2** exhibits a lower value $\text{EM}_{\text{thermo}}$ than **T1**—*i.e*. lower relative stability of the **T2** template duplex compared to the stability provided purely by non-cooperative binding between two sets of identical recognition motifs (K_a^{Ind}) within the ternary complex.

The fitting procedure employed in the extraction of kinetic parameters pertaining to the autocatalytic pathways for **T1** and **T2** was also employed in determining the kinetic parameters specific to the two crosscatalytic pathways. Fitting of kinetic data from the two crosscatalytic experiments, however, required addition of another kinetic parameter: association constant for the formation of the heteroduplex [**T1**·**T2**]. The template heteroduplex, comprising one molecule of each replicator, contains one more rotatable bond than [**T2**·**T2**] homoduplex and one rotatable bond less than the [**T1**·**T1**] homoduplex, and, therefore, the stability of this heteroduplex is bracketed by the stabilities of the two homoduplexes. Applying rotor increments as described[200] by Page and Jencks, the value for the [**T1**·**T2**] heteroduplex stability was estimated as $35.0 \times 10^6\ \text{M}^{-1}$. Utilising the calculated estimate of $K_\text{a}^{\text{Heteroduplex}}$ together with the kinetic parameters obtained thus far from the fitting of autocatalytic data in the kinetic models for each crosscatalytic system allowed determination of the recognition-mediated crosscatalytic rate constants for the formation of each *trans* product in the corresponding crosscatalytic template (**Table 3.2**).

The recognition-mediated rate constants and kinetic effective molarities (calculated with respect to the bimolecular rate constants), obtained through kinetic fitting of the crosscatalytic kinetic data, agree with the corresponding time course data for each replicator. Specifically, the *trans*-**T2** template was found to be significantly worse at

Table 3.2 Kinetic parameters, obtained using SimFit, for the crosscatalytic reactions between nitrone $\mathbf{N^F}$ and maleimide **M1** ([**M1**] = [$\mathbf{N^F}$] = 5 mM, [*trans*-**T2**] = 1 mM) and **M2** ([**M2**] = [$\mathbf{N^F}$] = 5 mM, [*trans*-**T1**] = 1 mM) at 5 °C, in $CDCl_3$).

	T1 being made on **T2** *trans*	**T2** being made on **T1** *trans*
k_{uni} / 10^{-4} s^{-1}	4.67	20.3
$EM_{kinetic}$ / M	4.21	18.3

catalysing the formation of *trans*-**T1**, than **T1** is itself. In fact, the $EM_{kinetic}$ for this crosscatalytic pathway is more than 20 × smaller than that determined for the autocatalytic pathway leading to the formation of *trans*-**T1**. Nevertheless, the **T2** template was found to catalyse the formation of **T1** replicator—once formed, however, the **T1** product is sequestered within the more stable heteroduplex [**T1**·**T2**], and not released as freely into solution as from the homoduplex [**T1**·**T1**].

In contrast, the *trans*-**T2** template is formed on *trans*-**T1** with efficiency similar to that on the autocatalytic *trans*-**T2** template. The kinetic effective molarity determined for this pathway, 18.3 M, is actually more than twice as high as the $EM_{kinetic}$ for the corresponding autocatalytic pathway (**T2** being made on **T2** template). It is possible, that as a result of the larger distance between the recognition sites within the open conformation of *trans*-**T1** than the distance in *trans*-**T2**, the *trans*-**T1** template is capable of pre-organising the reactive sites in **M2** and $\mathbf{N^F}$ in a co-conformation more favourable for their reaction. The calculated % enhancement for this reaction showed that the templating ability of *trans*-**T1** is higher at earlier reaction time points. Such observation can be rationalised in terms of the differences in duplex stabilities—at earlier time points, *trans*-**T1** is the dominant product, and, therefore, the **T2** product formed in the reaction is mostly present in the form of the heteroduplex [**T1**·**T2**], which is less stable than the [**T2**·**T2**] homoduplex, allowing higher catalytic turnover. Nevertheless, the affinity of *trans*-**T1** template for the components **M1** and $\mathbf{N^F}$ is lower than the affinity of **T2** for the same components, and as the **T2** continues to increase in concentration, the homoduplex [**T2**·**T2**] becomes the dominant species in solution, and the efficiency of the **T1**-making-**T2** pathway decreases.

3.4.2.6 Kinetic simulations

The kinetic parameters obtained through the kinetic fitting procedure for both replicators can be employed in kinetic simulations, allowing fast screening of numerous, potentially experimentally challenging reaction conditions. Throughout the template-instructed kinetic experiments reported here, the amount of template added at the beginning of the reaction was always maintained at 20 mol% (1 mM). By employing kinetic simulations,

the relevant extracted kinetic parameters can be directly employed in determining the kinetic behaviour of the **T1** and **T2** replicators in the presence of different amounts of mol% of preformed auto- and crosscatalytic templates. Systematic investigation of how these replicators perform under various instructed template conditions can reveal, for example, how much template is required by each replicator for successful initiation of the autocatalytic cycle or crosscatalytic cycle. To this end, a simulation file examining the formation of each replicator was run with increasing amounts of autocatalytic template (*i.e.* **T1** added to **T1**) added at $t = 0$ (0 mM, 0.125 mM, 0.25 mM, 0.5 mM, 1 mM, 2 mM). The simulated concentration *vs* time graphs and rate *vs* time graphs[d] are portrayed in **Figure 3.9**.

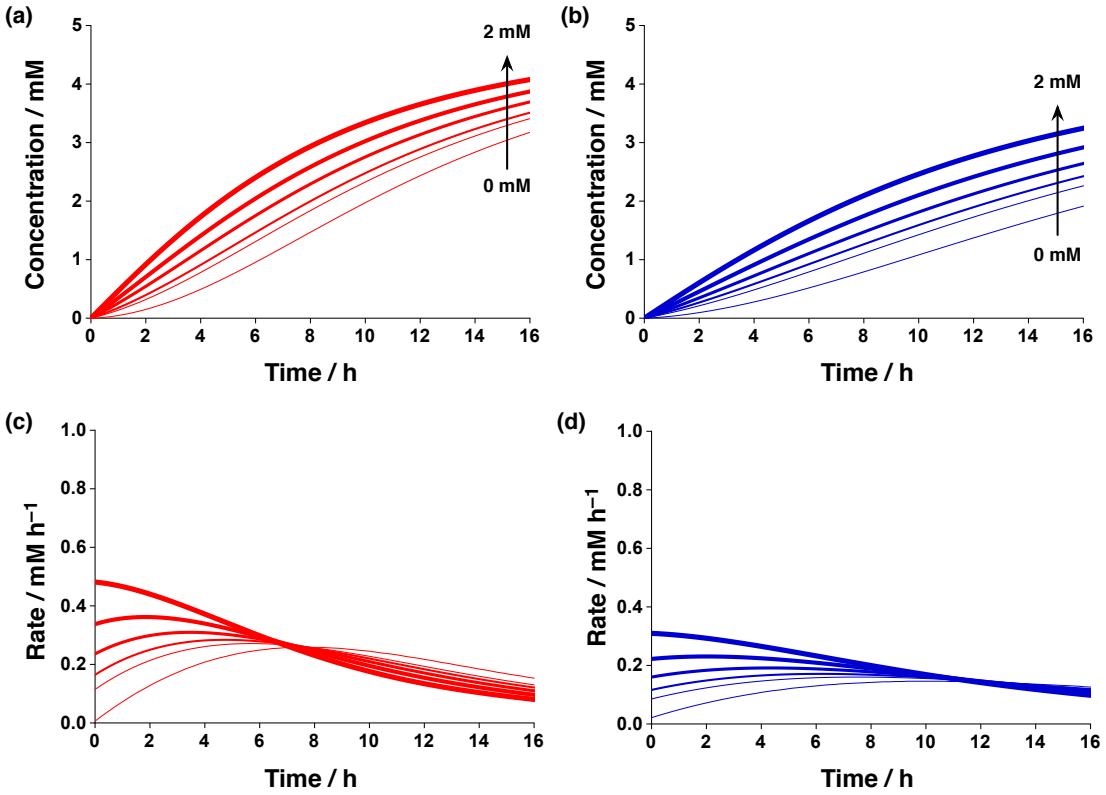


Figure 3.9 Simulated concentration *vs* time profile for the formation of **(a)** *trans*-**T1** in the presence of preformed *trans*-**T1** and **(b)** *trans*-**T2** in the presence of preformed *trans*-**T2**; simulated rate *vs* time profiles for the formation of **(c)** *trans*-**T1** in the presence of preformed *trans*-**T1** and **(d)** *trans*-**T2** in the presence of preformed *trans*-**T2**. [*trans*-**T1**] = [*trans*-**T2**] = 0, 0.125, 0.25, 0.5, 1 and 2 mM. Line thickness increases with increasing concentration of added preformed template (the thinnest line represents simulated data with no template added). Data is corrected for the template added.

The kinetic simulations (**Figure 3.9**) investigating the effect of preformed template on the reaction profile of each replicator within the autocatalytic pathway, show that in

[d]The rate (d[C]/dt) profile for each simulated condition was determined by computing the first derivative of a sixth order polynomial fitted to the simulated concentration *vs* time data.

order for the self-replicating reactions to proceed with the highest possible efficiency at $t = 0$, more than 20 mol% of preformed template (under the experimental conditions employed here) needs to be added to both **T1** and **T2**. The templating effect is particularly apparent from the rate *vs* time plots (**Figure 3.9c** and **d**), which highlight that a slightly higher amount of template (between 1 mM and 2 mM) is needed for **T1** and **T2** replicators to be able to operate at maximum replication efficiency right from $t = 0$, than employed in the kinetic experiments. Nevertheless, the addition of even as little as 2 mol% (0.125 mM) of preformed template results in a noticeable shortening of the lag period for each replicator.

The rate of reaction in self-replicating systems depends on the formation of the ternary, catalytically-active complex. In turn, the quantity of this complex within a reaction mixture is governed by the concentration of the replicator template and the building blocks, as well as the association strength of the individual binding event (K_a^{Ind}). Using the kinetic simulations of the autocatalytic **T1** and **T2** pathways described thus far, it was possible to also calculate the concentration of the ternary complex [**M1**·**N**$^{\mathbf{F}}$·**T1**] (**Figure 3.10a**) and [**M2**·**N**$^{\mathbf{F}}$·**T2**] (**Figure 3.10b**) within each pathway, in the presence of varying amount of preformed autocatalytic template.

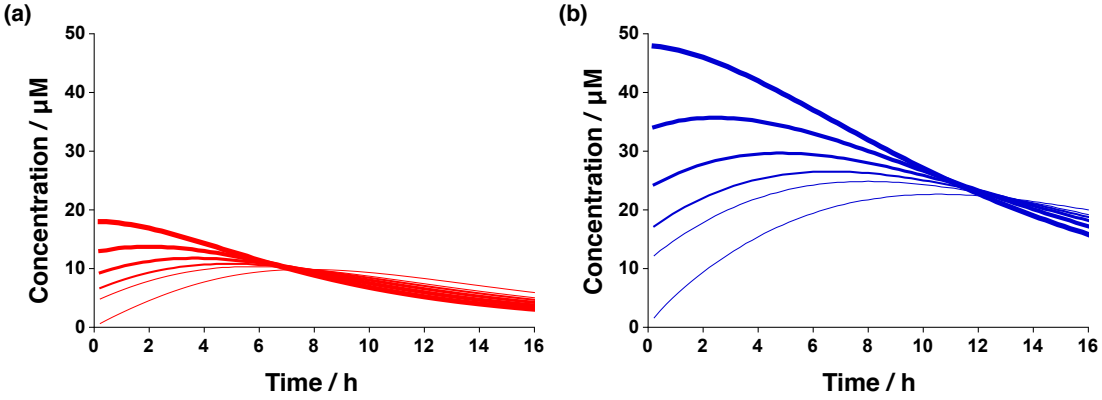


Figure 3.10 Simulated concentration *vs* time profile for **(a)** ternary complex [**M1**·**N**$^{\mathbf{F}}$·**T1**] formed in the reaction between **M1** with **N**$^{\mathbf{F}}$ in the presence of preformed *trans*-**T1** and **(b)** ternary complex [**M2**·**N**$^{\mathbf{F}}$·**T2**] formed in the reaction between **M2** with **N**$^{\mathbf{F}}$ in the presence of preformed *trans*-**T2** ([*trans*-**T1**] = [*trans*-**T2**] = 0, 0.125, 0.25, 0.5, 1 and 2 mM). Line thickness widens with increasing concentration of added preformed template (the thinnest line represents simulated data with no template added).

As can be expected based on the relationship between the concentration of the catalytically-active ternary complex and reaction rate, the peak formation of product (*i.e.* maximum reaction rate) is observed at the time when the concentration of [**M1**·**N**$^{\mathbf{F}}$·**T1**] and [**M2**·**N**$^{\mathbf{F}}$·**T2**] complexes is the highest in the reaction mixture. The time at which the maximum rate and the concentration of ternary complexes is observed depends on the amount of preformed template added—addition of 2 mM of preformed template

produces the highest amount of ternary complexes for both **T1** and **T2**. Interestingly, the **T2** replicator is more efficient at assembling the unreacted components in to a catalytically-active complex—in each simulated condition, the concentration of the ternary complex is more than twice is high as that calculated for **T1** under analogous conditions. The higher proportion of ternary complex in the **T2** autocatalytic pathway reflects the higher value of K_a^{Ind} for the association of **M2** with $\mathbf{N^F}$ than the corresponding value for interaction between **M1** with $\mathbf{N^F}$. Despite the advantage afforded to **T2** by the higher K_a^{Ind}, it is less efficient at templating its own formation than **T1**, as evidenced by the rate *vs* time profiles for these replicators.

The extracted kinetic parameters can also be employed to simulate the kinetic behaviour of the **T1** replicator in the presence of different concentrations of *trans*-**T2** template. Kinetic simulations of this crosscatalytic pathway, that was shown to operate inefficiently through the kinetic analysis, can facilitate determination of the effect of template concentration on the ability of *trans*-**T2** to template the formation of *trans*-**T1**. The simulation output portrayed as a concentration *vs* time plot (**Figure 3.11a**) facilitated comparison of the *trans*-**T1** reaction profile in the presence of 0, 1, 5 and 10 mM of *trans*-**T2**.

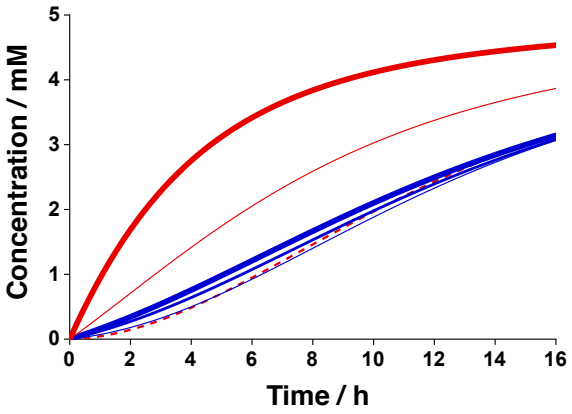


Figure 3.11 Simulated concentration *vs* time profile for the formation of *trans*-**T1** in the presence of preformed *trans*-**T2**. [*trans*-**T2**] = 0 (dashed red line), 1, 5 and 10 mM. The thinnest blue line represent simulated data with 1 mM template added and progressive thickening of the lines indicates increasing concentrations of template added. Formation of *trans*-**T1** in the presence of 1 mM (thin red line) and 10 mM (thick red line) of *trans*-**T1** is shown for comparison.

These results showed that, while at the experimentally examined *trans*-**T2** concentration of 1 mM, there is virtually no change in the time course of *trans*-**T1** formation (particularly in the length of the lag period), the addition of *trans*-**T2** at a significantly higher concentration, 5 and 10 mM, begins to exert a more noticeable effect on the reaction profile of **T1**—as evidenced by the slight decrease in the length of the lag period. The decrease is not as marked as that observed in the reaction instructed with as

little as 1 mM of *trans*-**T1** template. These simulations suggests that *trans*-**T2** is not a suitable template for the formation of *trans*-**T1**, but can exert a catalytic effect at high template concentrations.

Analysis of the of the K_a^{Duplex} values for the two homo duplexes and the heteroduplex revealed that **T2** is capable of forming significantly stronger complexes than **T1**. Specifically, the stability exhibited by the duplexes in the **T1**–**T2** network follows the order of: [**T1**·**T1**] < [**T1**·**T2**] < [**T2**·**T2**]. This difference in the stability of product duplexes is also the reason behind the lack of apparent catalytic effect of **T2** template on the formation of **T1**. The presence of **T2** replicator catalyses the formation of **T1** with a low $EM_{kinetic} = 4.21$ M. As a result of the higher stability of the heteroduplex [**T1**·**T2**] than the homoduplex [**T1**·**T1**], the **T1** formed in the reaction is sequestered within the product complexes, not available for reaction. In order to further understand the influence of the template duplex stability on the distribution of **T1** within the mixture, the formation of [**T1**·**T1**], [**T1**·**T2**] and [**T2**·**T2**] in a condition where the reagents required to make **T1** are instructed with 20 mol% of preformed **T2**, was plotted (**Figure 3.12a**).

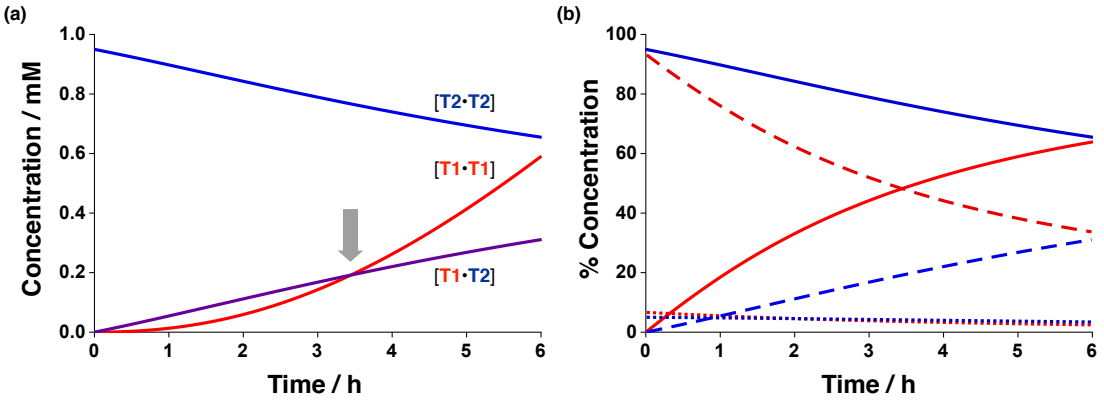


Figure 3.12 **(a)** Simulated concentration of [**T1**·**T1**], [**T1**·**T2**] and [**T2**·**T2**] in a reaction of **M1** and $\mathbf{N}^{\mathbf{F}}$ in the presence of 20 mol% of preformed *trans*-**T2**. Concentrations of homoduplexes have been corrected for the presence of two template molecules within each duplex. The grey arrow indicates the crossover point where the concentration of [**T1**·**T1**] exceeds the concentration of template **T1** bound within heteroduplex [**T1**·**T2**]. **(b)** Percentage distribution of the preformed added template *trans*-**T2** (1 mM) and the replicator *trans*-**T1** formed within the reaction (at a particular time) within a homoduplex (full line), heteroduplex (dashed line) and within other recognition-mediated complexes and in unbound state (dotted line); calculated distribution is shown in red for **T1** and in blue for **T2**.

The simulation outcome clearly shows that there is very little of the [**T1**·**T1**] duplex present at the initial stage of the reaction (*ca.* up to 4 hours) and instead the homoduplex [**T2**·**T2**] and heteroduplex [**T1**·**T2**] are the dominant products in solution. As a result of the significant proportion of **T1** bound within the more stable heteroduplex, the **T1** template is less available in its free form, necessary for the formation of **T1** replicator

through the catalytically most efficient pathway mediated by the ternary complex [**M1**·**N**$^{\mathbf{F}}$·**T1**].

Probing the fate of both the preformed instructing template **T2** and template **T1** formed in the reaction further, the kinetic simulations were used to calculate[e] the % of each template bound within a homoduplex, heteroduplex as well as the combined concentration of **T1** and **T2** present in other recognition-mediated complexes and in its free form (**Figure 3.12b**). The simulation focused on the initial six hours in the reaction time course—a time period encompassing the crossover point where the concentration of homoduplex [**T1**·**T1**] surpasses the concentration of the heteroduplex. Within this period of six hours, the concentration of both templates remains within 1 mM. **Figure 3.12b** shows that both **T1** and **T2** are predominantly present in the form of homo- and heteroduplexes—less than 10% of each template is present in other recognition-mediated complexes or in its free form. This small fraction of template includes also the ternary catalytically-active complexes—the non-covalently bound species that are actually present at fairly low concentrations (*ca.* 10–50 µM range) at any one time within the reaction mixture.

The variation in stability of template duplexes was also suggested to play a role in the observed % enhancement observed the **T1**-being made on **T2**-crosscatalytic pathway at early and late reaction times. More specifically, it was proposed in the discussion of this crosscatalytic pathway that the templating ability of *trans*-**T1** is higher at earlier reaction time points as a result of a higher catalytic turnover of both products, made possible by the dominance of the [**T1**·**T2**] heteroduplex, as opposed to the stronger [**T2**·**T2**] homoduplex, in the mixture. Using kinetic simulations and the obtained reaction and recognition parameters, it was possible to probe this phenomenon by examining the distribution of **T1** and **T2** templates in the various duplexes in a reaction of **M2** and **N**$^{\mathbf{F}}$ in the presence of 20 mol% of preformed **T1** (**Figure 3.13a**). The kinetic simulation confirmed that up to *ca.* $t = 2.5$ hours, the concentration of [**T1**·**T2**] exceeds the concentration of [**T2**·**T2**]. Clearly, the heteroduplex is present at a higher concentration than the **T2** homoduplex only temporarily—an observation that explains why **T1** exerts a stronger catalytic effect on the formation of **T2** early in the reaction than at later in the time course. Analysis of the calculated percentages of both the instructing template and the **T2** formed in the reaction (**Figure 3.13b**) showed that again, both templates are predominantly present in the form of product duplexes. The alternative visualisation of the product distribution in **Figure 3.13b** illustrates the gradual decrease in the concentration of [**T1**·**T1**] and the fraction of **T2** within

[e]The percentage distribution for the template formed in the reaction was calculated relative to the overall concentration of the said template present in the mixture at a particular time. For the preformed template, the distribution was calculated relative to the concentration added at $t = 0$ hours.

[**T1**·**T2**] and the simultaneous increase in the proportion of [**T2**·**T2**] and **T1** bound within [**T1**·**T2**]. This decrease in the proportion of **T1** in [**T1**·**T2**], as a result of the concentration of **T2** increasing in the reaction mixture over time is again reflected directly in the observation that **T1** template can catalyse the formation of **T2** more efficiently at earlier reaction times.

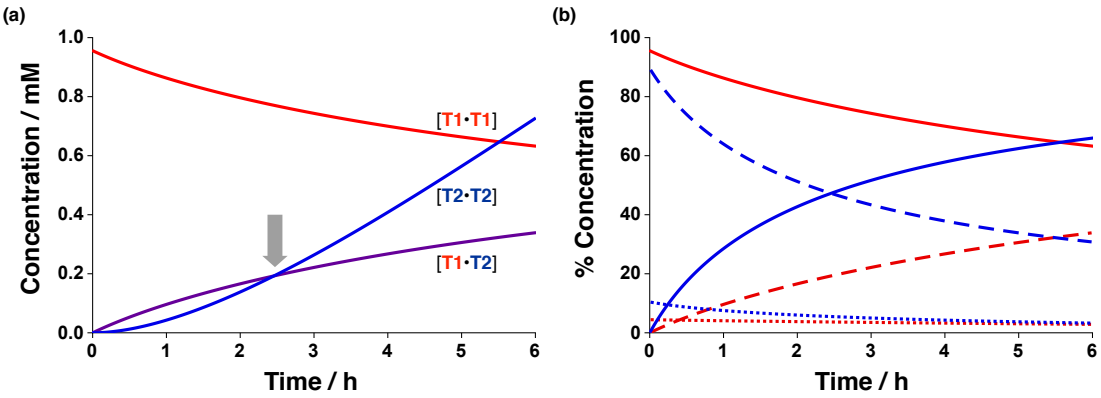


Figure 3.13 **(a)** Simulated concentration of [**T1**·**T1**], [**T1**·**T2**] and [**T2**·**T2**] in a reaction of **M2** and $\mathbf{N^F}$ in the presence of 20 mol% of preformed *trans*-**T1**. Concentrations of homoduplexes have been corrected for the presence of two template molecules within each duplex. The grey arrow indicates the crossover point where the concentration of [**T2**·**T2**] exceeds the concentration of template **T2** bound within heteroduplex [**T1**·**T2**]. **(b)** Percentage distribution of the preformed added template *trans*-**T1** (1 mM) and the replicator *trans*-**T2** formed within the reaction (at a particular time) within a homoduplex (full line), heteroduplex (dashed line) and within other recognition-mediated complexes and in unbound state (dotted line); calculated distribution is shown in red for **T1** and in blue for **T2**.

3.5 Competition kinetic experiments

Kinetic analyses of the two replicators *trans*-**T1** and *trans*-**T2** in isolation showed that *trans*-**T1** template has the capacity to template its own formation as well as cross-catalyse the formation of *trans*-**T2**, while the *trans*-**T2** template was found to be only efficient as an autocatalyst. With these results at hand, the kinetic behaviour of **T1** and **T2** was examined under experimental conditions enforcing competition for the shared building block, nitrone $\mathbf{N^F}$. The possibility of selectively directing the system containing all three building blocks, **M1**, **M2** and $\mathbf{N^F}$, where both replicators are produced simultaneously, towards selective formation of a single product, *trans*-**T1** or *trans*-**T2**, within the competition environment was assessed through kinetic experiments employing addition of instructing, preformed template (added at $t = 0$ hour).

Four experimental, instructing conditions were designed in order to fully probe the possibility of using template to direct the outcome of the competition within the network

of two replicators, **T1**–**T2**. In each experiment, an equimolar solution of nitrone (the shared, limiting building block) and both maleimides ([$\mathbf{N^F}$] = [**M1**] = [**M2**] = 5 mM) was prepared in $CDCl_3$. In the uninstructed kinetic experiment the product evolution was examined 5 °C in the absence of any added template at regular intervals by 470.3 MHz $^{19}F\{^1H\}$ NMR spectroscopy. The instructed experiments examined the formation of **T1** and **T2** in the presence of 20 mol% preformed (i) *trans*-**T1**, (ii) *trans*-**T2** or (iii) *trans*-**T1** and *trans*-**T2** simultaneously. Following mixing of the reaction components, the progress of each reaction was monitored regularly over the period of 16 hours. Deconvolution of the obtained kinetic data enabled construction of concentration *vs* time plots for each of the competition experiments (**Figure 3.14**).

Kinetic analyses of the **T1**–**T2** network under competition conditions further highlighted the potential of small structural changes to exert a dramatic effect on a replicator's ability to compete for a limited, shared resource. Specifically, the uninstructed competition experiment (**Figure 3.14**) revealed that, as a result of the interplay of the recognition and reaction processes in the four catalytic cycles in the network, the normally less efficient template, *trans*-**T2**, is capable of reaching higher concentration than the replicator found to be more efficient in isolation, *trans*-**T1** ([*trans*-**T2**]/[*trans*-**T1**] ratio of 1.16 after 16 hours). Comparison of the [*trans*-**T2**]/[*trans*-**T1**] ratio formed after 16 hours in the three template-instructed experiments to the ratio determined in the uninstructed kinetics showed that the system can be successfully directed to make the added instructing template with increased selectivity. In particular, addition of *trans*-**T1** directed the **T1**–**T2** network towards increased formation of *trans*-**T1** ([*trans*-**T2**]/[*trans*-**T1**] = 0.81), while addition of preformed *trans*-**T2** further biased the ratio in the favour of *trans*-**T2** ([*trans*-**T2**]/[*trans*-**T1**] = 1.58). Simultaneous addition of both templates enabled *trans*-**T1** replicator to outcompete *trans*-**T2**, resulting in a [*trans*-**T2**]/[*trans*-**T1**] ratio of 0.91 after 16 hours.

Interestingly, when the [*trans*-**T2**]/[*trans*-**T1**] ratio is examined at t = 4 hours, an increase in selectivity for **T2** replicator is observed in the condition with no template and in the presence of *trans*-**T2**, affording values of 1.64 and 2.42, respectively. The ratio of [*trans*-**T2**]/[*trans*-**T1**] at the same time point, in the kinetic experiment instructed with *trans*-**T1** and both templates simultaneously remained virtually unchanged. The absence of changes in the replicator ratios in these two conditions can be most likely explained by the fact that *trans*-**T1** is a catalyst for the formation of both products, whereas *trans*-**T2** only catalyses the formation of itself. Therefore, at earlier reaction time points, formation of *trans*-**T2** is being dramatically enhanced by the presence of *trans*-**T2** template, whilst the formation of *trans*-**T1** relies on the template *trans*-**T1** produced within the reaction mixture.

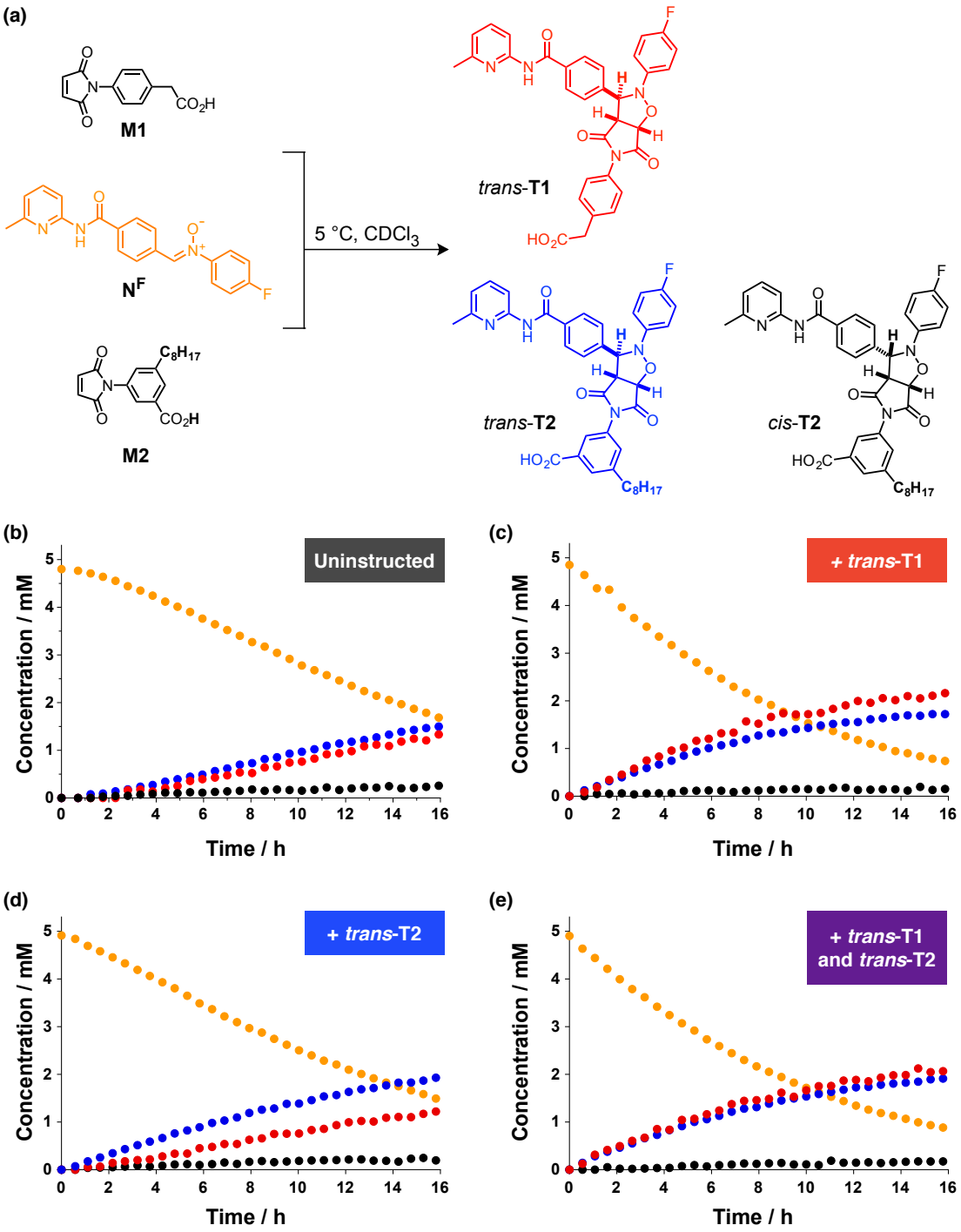
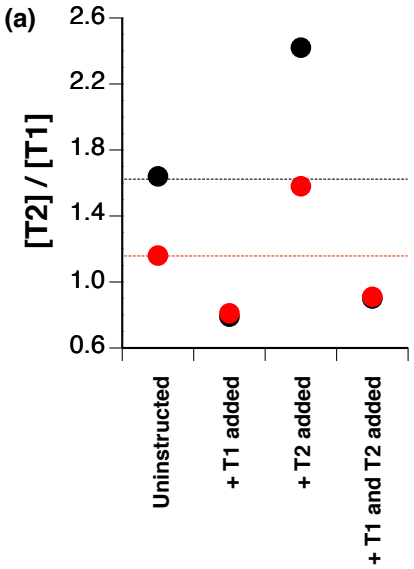


Figure 3.14 Concentration *vs* time profiles for reaction of nitrone $\mathbf{N^F}$ with **M1** and **M2** (all components at 5 mM) in the presence of **(a)** no template, **(b)** 20 mol% *trans*-**T1**, **(c)** 20 mol% *trans*-**T2** and **(d)** 20 mol% of both *trans*-**T1** and *trans*-**T2**. Experimental data is represented as circles. $\mathbf{N^F}$ is shown in orange, *trans*-**T1** in red, *trans*-**T2** in blue and *cis*-**T2** in black. Formation of *cis*-**T1** was not observed. All reactions were carried out in $CDCl_3$ at 5 °C and the reaction progress was monitored by $^{19}F\{^1H\}$ NMR spectroscopy. Concentrations were determined relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard. Data is corrected for the template added.

The same increase is not observed in the kinetic experiments instructed by both templates, indicating that the increase in the selectivity for the formation of **T2** replicator is countered by the presence of *trans*-**T1** template within the mixture, which enhances the formation of **T1** more efficiently than **T2**.

Comparison of the [*trans*-**T2**]/[*trans*-**T1**] ratio formed under the four competition conditions (**Figure 3.15a**) reveals that the behaviour of the **T1**–**T2** replicating system can be likened to that of an IF molecular logic gate. IF type of Boolean logic operation specifies that a fulfilment of a specific condition, results in another condition being true. In this molecular logic gate, the behaviour corresponds to the addition of preformed template **T1** and the conditional response is the decrease in the [*trans*-**T2**]/[*trans*-**T1**] ratio formed under each condition (**Figure 3.15b**).



(b)

IF Logic Operation

T1	T2	[T2] / [T1]↓
0	0	0
1	0	1
0	1	0
1	1	1

Figure 3.15 **(a)** The effect of instructing the **T1**–**T2** replicating network with preformed template (added at t = 0), on the ratio of [*trans*-**T2**]/[*trans*-**T1**] formed after four hours (black circles) and 16 hours (red circles). Dashed lines represent the [*trans*-**T2**]/[*trans*-**T1**] ratio determined in the uninstructed kinetic experiment, for comparison. **(b)** IF logic gate describes the behaviour of the system in response to template **T1** added, where 1 corresponds to presence of template as well as a decrease in the [*trans*-**T2**]/[*trans*-**T1**] ratio).

In order to facilitate comparison of the outcomes across the various experimental conditions used to examine the behaviour of **T1** and **T2** replicators in isolated and competition experiments in this chapter, an overview of the results from each experiment, highlighting the difference between instructed and uninstructed conditions is provided in **Figure 3.16**.

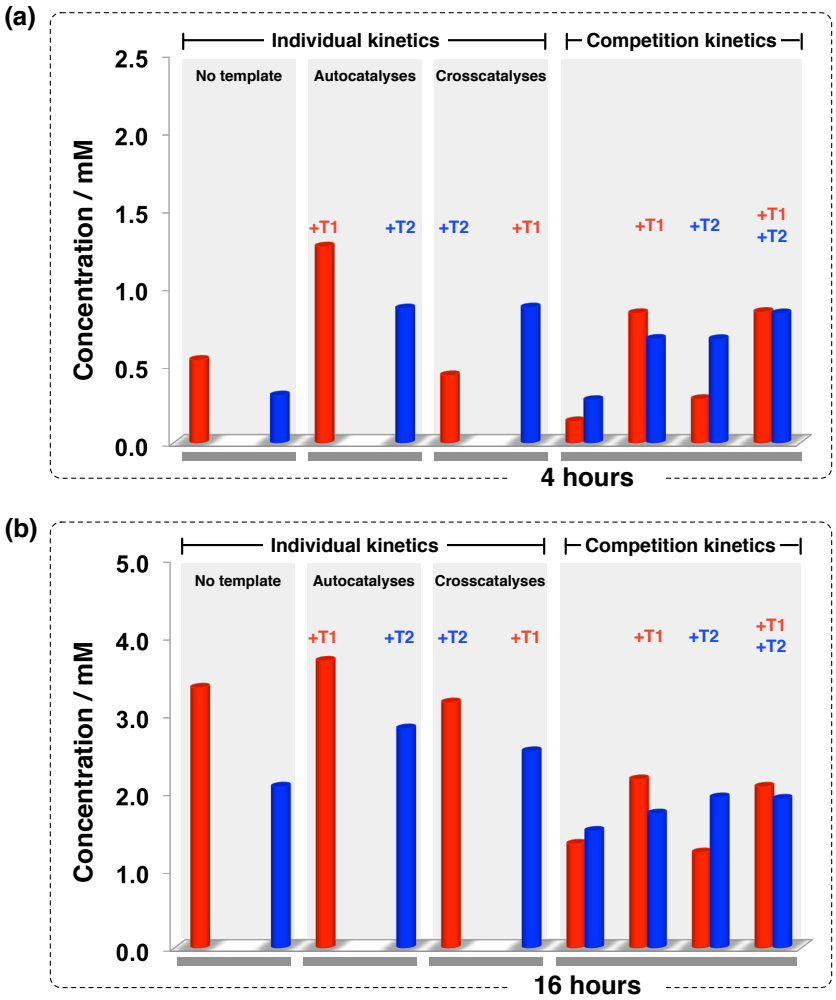


Figure 3.16 Overview of results obtained from the kinetic experiments carried out in order to analyse the replicating network after **(a)** four hours and **(b)** 16 hours. Left section represents results determined for each replicator in isolation whilst the right hand side showed outcome of competition kinetics ($CDCl_3$, at 5 °C). **T1** is shown in red and **T2** is represented in blue. The concentrations templates formed have been corrected for the template added.

3.5.1 Stability of the replicator duplexes: an NMR study

The concentration *vs* time plots for the competition experiments, examining the simultaneous formation of **T1** and **T2** replicators, were constructed using kinetic data determined by $^{19}F\{^1H\}$ NMR spectroscopy. Throughout the competition experiments, the behaviour of the network was also monitored by ^1H NMR spectroscopy. Examination of the ^1H NMR data for these kinetic experiments revealed interesting pattern in the chemical shifts of the resonances arising from **T1** and **T2** replicators, in particular the resonances associated with the *trans* proton on the isoxazolidine ring (**Figure 3.17**). These changes can be illustrated most clearly on the example of the competition kinetic experiment instructed with 20 mol% of *trans*-**T1** (**Figure 3.17**). At the beginning of the reaction, **T1** replicator, added at concentration of 1 mM is the dominant species in

solution. In contrast, no **T2** template is present at the beginning of the reaction. The competition experiment contains the building blocks required for the synthesis of both replicators, and in the presence of preformed **T1** template, both **T1** and **T2** increase in concentration quickly. The increase in the concentration of **T2** replicator in the reaction mixture over time is associated with a change in the chemical shift of the resonance associated with the *trans* proton present in the **T1** replicator, but not the resonance of **T2** replicator itself. The variation in the chemical shift, or lack of it for replicator **T2**, demonstrates quite strikingly the ability of **T1** and **T2** replicators to associate in both homo- ([**T1**·**T1**] and [**T2**·**T2**]) and heteroduplexes ([**T1**·**T2**]), giving rise to a chemical shift that is dependent on the mole fraction of the template bound within the homoduplex and the heteroduplex.

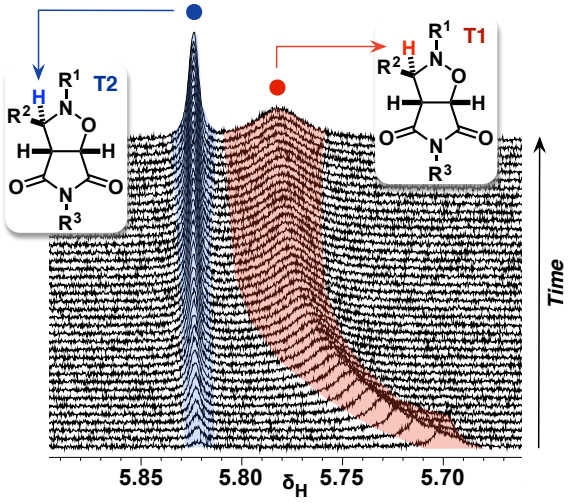


Figure 3.17 Partial plot of arrayed 499.9 MHz ^{1}H NMR spectra showing the changes in the resonances associated with the *trans* protons on the isoxazolidine ring of **T1** (red) and **T2** (blue) replicators formed within a competition experiment instructed with **T1** template ([**M1**] = [**M2**] = [$\mathbf{N^F}$] = 5 mM, [*trans*-**T1**] = 1 mM, $CDCl_3$, 5 °C) over time.

In order to develop a better understanding of the variation in the chemical shifts for the resonances associated with both replicators, a more controlled set of experiments was designed. To this end, **T1** and **T2** replicators were examined under five different concentration conditions. Specifically, each of the five conditions contained a solution of these two replicators at a combined concentration of 10 mM, in $CDCl_3$ (for details, see **Chapter 9**), beginning with (i) 100% of **T1** and progressively increasing the fraction of **T2**: (ii) 25%, (iii) 50%, (iv) 75% and finally (v) 100% of **T2**. Each mixture was analysed by ^{1}H NMR (**Figure 3.18a**) and ^{19}F{^{1}H} NMR spectroscopy **Figure 3.18b**). These five conditions allowed the changes in the chemical shifts arising from the resonances of both replicators to be examined in conditions where (i) only homoduplex [**T1**·**T1**] and (v) only homoduplex [**T2**·**T2**] is present, as well as three conditions (ii to iv) in between examining the mixtures where both homo and heteroduplex are present.

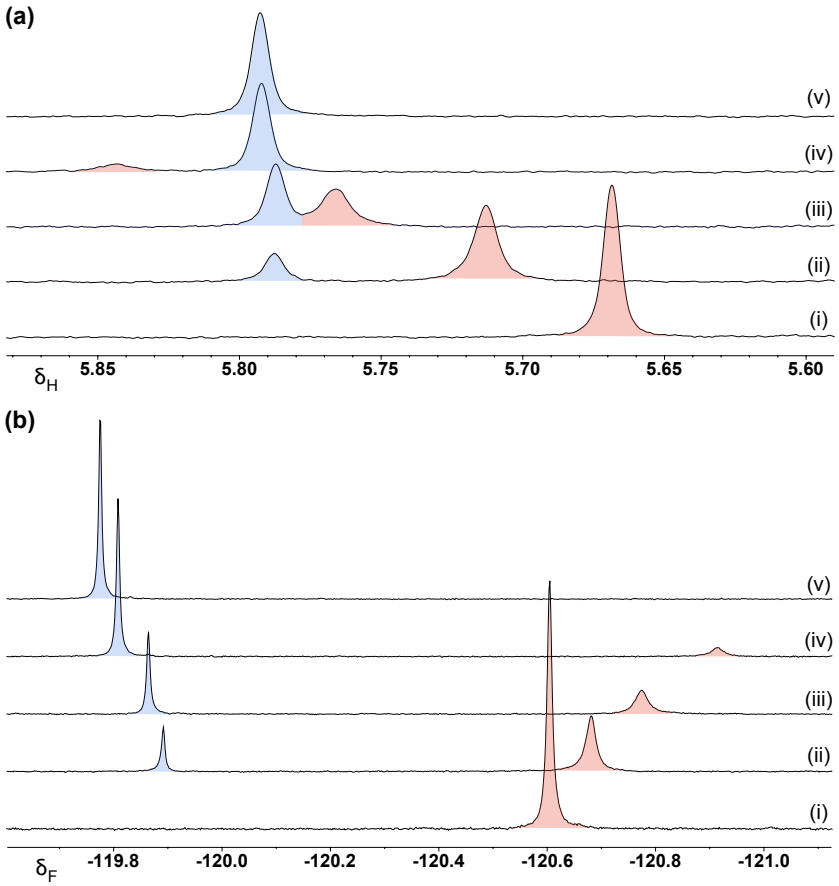


Figure 3.18 Overlay of partial **(a)** 400.1 MHz ^{1}H NMR spectra (RT) and **(b)** 376.5 MHz ^{19}F$\{^{1}$H$\}$ NMR spectra (RT) showing the *trans* resonances belonging to the **T1** and **T2** cycloadducts. The chemical shift changes for these *trans* products conditions were examined at five different conditions: starting with (i) **T1** only and progressively increasing the content of **T2** to (ii) 25%, (iii) 50%, (iv) 75% and finally (v) 100%. The combined concentration of the **T1** and **T2** templates was always kept constant at 10 mM.

The ^{1}H NMR chemical shifts arising from the *trans* resonances of **T1** in [**T1**·**T1**] duplex (δ_{H} 5.67) and **T2** in [**T2**·**T2**] duplex (δ_{H} 5.79) were extracted from the ^{1}H NMR data and the chemical shift for each template in the heteroduplex was estimated (**Table 3.3**). Formation of each product duplex (**Table 3.3**) is associated with a particular $K_{\mathrm{a}}{}^{\mathrm{Duplex}}$. The $K_{\mathrm{a}}{}^{\mathrm{Duplex}}$ for the formation of [**T1**·**T1**] and [**T2**·**T2**] homoduplexes were extracted through the fitting of kinetic data, whereas the value for the [**T1**·**T2**] was estimated according to the procedure described in the kinetic fitting section earlier on in this chapter.

Using the chemical shift values and the corresponding association constants for each homo and heteroduplex, summarised in **Table 3.3**, it was possible to simulate the steady state distribution of **T1** and **T2** replicators within the various product duplexes present

in the system. Nine different concentration conditions were examined using Gepasi 3 simulation software, starting with **T1** at 1 mM and **T2** at 9 mM (overall 10 mM), and progressively reversing the ratio in 1 mM increments.

Table 3.3 Homo- and heteroduplex association constant values and the corresponding ^{1}H NMR chemical shift values (ppm) associated with the *trans* resonances for **T1** and **T2** replicators.

	[T1•T1]	[T2•T2]	[T2•T1] (T1)	[T2•T1] (T2)
K_a^{Duplex} / 10^6 M^{-1}	17.88	68.60	35.0*	35.0*
^{1}H NMR (400.1 MHz) Chemical shift / ppm	5.67	5.79	5.91*	5.77*

* Denotes values that were estimated

Normally, the simulated steady state distributions of **T1** and **T2** templates would be used in **Equation 3.5** and **Equation 3.6** to calculate the predicted observed chemical shift for each template.

$$\delta_{\mathbf{T1}}\ (\text{observed}) = f_{[\mathbf{T1}]} \cdot \delta_{[\mathbf{T1}]} + f_{[\mathbf{T1}\cdot\mathbf{T1}]} \cdot \delta_{[\mathbf{T1}\cdot\mathbf{T1}]} + f_{[\mathbf{T1}\cdot\mathbf{T2}]} \cdot \delta_{[\mathbf{T1}\cdot\mathbf{T2}]} \qquad (\mathbf{3.5})$$

$$\delta_{\mathbf{T2}}\ (\text{observed}) = f_{[\mathbf{T2}]} \cdot \delta_{[\mathbf{T2}]} + f_{[\mathbf{T2}\cdot\mathbf{T2}]} \cdot \delta_{[\mathbf{T2}\cdot\mathbf{T2}]} + f_{[\mathbf{T1}\cdot\mathbf{T2}]} \cdot \delta_{[\mathbf{T1}\cdot\mathbf{T2}]} \qquad (\mathbf{3.6})$$

However, as a result of the relatively high values for K_a^{Duplex}, there is virtually no **T1** or **T2** template present in the free, uncomplexed state at the examined concentrations. This means that the terms representing the fraction of **T1** and **T2** in its free form can be neglected. Therefore, **Equation 3.5** and **Equation 3.6** can be simplified to give the final **Equation 3.7** for **T1** and **Equation 3.8** for **T2** as shown below.

$$\delta_{\mathbf{T1}}\ (\text{observed}) = f_{[\mathbf{T1}\cdot\mathbf{T1}]} \cdot \delta_{[\mathbf{T1}\cdot\mathbf{T1}]} + f_{[\mathbf{T1}\cdot\mathbf{T2}]} \cdot \delta_{[\mathbf{T1}\cdot\mathbf{T2}]} \qquad (\mathbf{3.7})$$

$$\delta_{\mathbf{T2}}\ (\text{observed}) = f_{[\mathbf{T2}\cdot\mathbf{T2}]} \cdot \delta_{[\mathbf{T2}\cdot\mathbf{T2}]} + f_{[\mathbf{T1}\cdot\mathbf{T2}]} \cdot \delta_{[\mathbf{T1}\cdot\mathbf{T2}]} \qquad (\mathbf{3.8})$$

Figure 3.19 shows the calculated predicted chemical shifts corresponding to the *trans* resonances for **T1** and **T2**, plotted against the concentration of **T1** present in the system, where the combined concentration of **T1** and **T2** is always equal to 10 mM).

Comparison of the calculated chemical shifts, determined by simulating the distribution of **T1** and **T2** within the product duplexes, to the experimentally determined ^{1}H NMR data shows a clear qualitative agreement. The high stability of the [**T2**·**T2**] is reflected in the lack of changes in the chemical shift of the resonance associated with this product, which instead remains virtually unchanged throughout the experiments. The opposite is true for the [**T1**·**T1**]—the K_a^{Duplex} for this product dimer compared to the other duplexes is the lowest, resulting in quite a dramatic change in the chemical shift of the resonance arising from the **T1** template (almost 0.2 ppm). The NMR analysis of the heteroduplex formation shows that the association constants determined by kinetic

fitting can be used successfully to predict the behaviour of the chemical shifts for the two catalytic templates in the highly interconnected network.

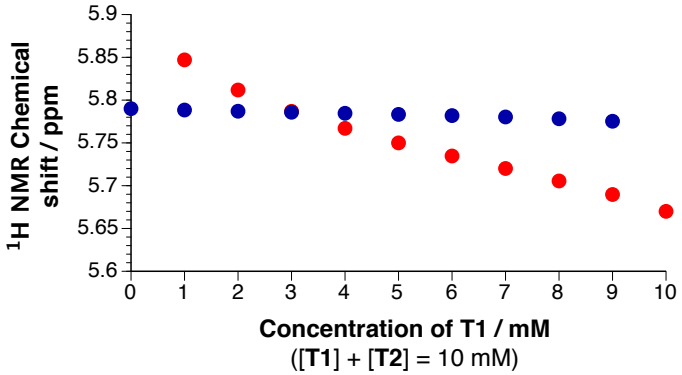


Figure 3.19 Simulated changes in chemical shift for the isoxazolidine proton in **T1** (red circles) and in **T2** (blue circles) observed as a result of the distribution of these products in homo and heteroduplexes. Data was calculated in Gepasi 3, using duplex association constants determined through fitting of kinetic data and estimated/determined chemical shifts for each duplex in the system.

3.5.2 Kinetic simulations of a two-replicator reaction network

Outcome of the isolated, non-competition kinetic experiments confirmed that both the *trans*-**T1** and the *trans*-**T2** template are capable of self-replication. The instructed non-competition experiments revealed that *trans*-**T1** is a more efficient and more diastereoselective replicator than *trans*-**T2**. Nevertheless, competition experiments, where the formation of **T1** and **T2** took place simultaneously, demonstrated that replicator *trans*-**T2**, while less efficient in isolation, outperforms *trans*-**T1** replicator in the competition for the shared nitrone resource as a result of the interplay of the two active autocatalytic cycles and the the single efficient crosscatalytic cycle (*trans*-**T1** does not form efficiently on *trans*-**T2** template). In order to probe the possibility of extending the kinetic simulations to a full kinetic model, incorporating the formation of both replicators simultaneously, the parameters for the **T1**–**T2** network were added to a kinetic model and the behaviour of the system was simulated under conditions identical to those employed in the uninstructed experimental kinetic experiment (**Figure 3.20**).

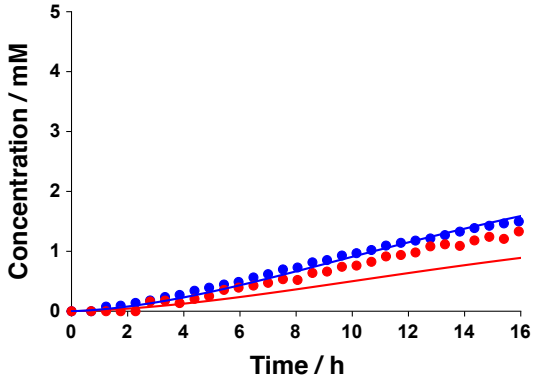


Figure 3.20 Concentration *vs* time profiles for reaction of nitrone $\mathbf{N^F}$ with **M1** and **M2** (all components at 5 mM) in the absence of template. Experimental data are represented as circles and simulated data are shown as full lines, where *trans*-**T1** is denoted in red and *trans*-**T2** in blue.

Comparison of the simulation outcome to the kinetic data obtained for **T1** and **T2** replicators under competition conditions showed only a partial agreement between the two sets of data (simulated and experimental). While the concentration *vs* time profile for the **T2** replicator approximated the experimental data very well, the simulation underestimated quite noticeably the formation of **T1** replicator. In this case, the kinetic and thermodynamic parameters determined through fitting of kinetic data for individual replicators in isolation did not sufficiently account for the kinetic behaviour of the network as a whole—most likely as a result of the interplay between the various reaction and recognition elements in the competition scenario that are not present in the isolated kinetic experiments. Future work on simulating networks of replicators under competition conditions will need to carefully consider and evaluate the implemented kinetic models for these networks and probe any potentially missing interactions. Despite these limitations, the simulated competition network does provide a good starting point for probing the behaviour of the **T1**–**T2** network, and will be discussed further in **Chapter 4**.

3.6 Summary

Two replicators, **T1** and **T2**, emerge through 1,3-dipolar cycloaddition reactions of two structurally-similar maleimide building blocks with a shared nitrone component (**Figure 3.21a**). Kinetic analyses, examining the formation of each replicator from its constituent components, both in the absence and presence of preformed template, revealed that even a small structural change in the position of one recognition site, here the identity of the –COOH group, results in the formation of two replicators with appreciably different catalytic abilities. In the absence of template, both **T1** and **T2**

replicators formed efficiently and with high diastereoselectivity for the catalytically-active *trans* cycloadduct, exhibiting a significant rate enhancement relative to the analogous kinetic experiments employing recognition-disabled maleimides. The ability of both templates to self-replicate was demonstrated successfully by the marked decrease in the length of each lag period upon addition of the respective autocatalytic template, nevertheless, replicator **T1** was found to be significantly more efficient at templating its own formation than **T2**. Kinetic fitting of the experimental data revealed that while **T2** replicator is formed faster through the bimolecular pathway and forms stronger interactions between the two complementary recognition units than replicator **T1**—*i.e.* it is capable of assembling the reaction components in to the catalytically-active ternary complex more efficiently—it also suffers from more severe product inhibition than replicator **T1**.

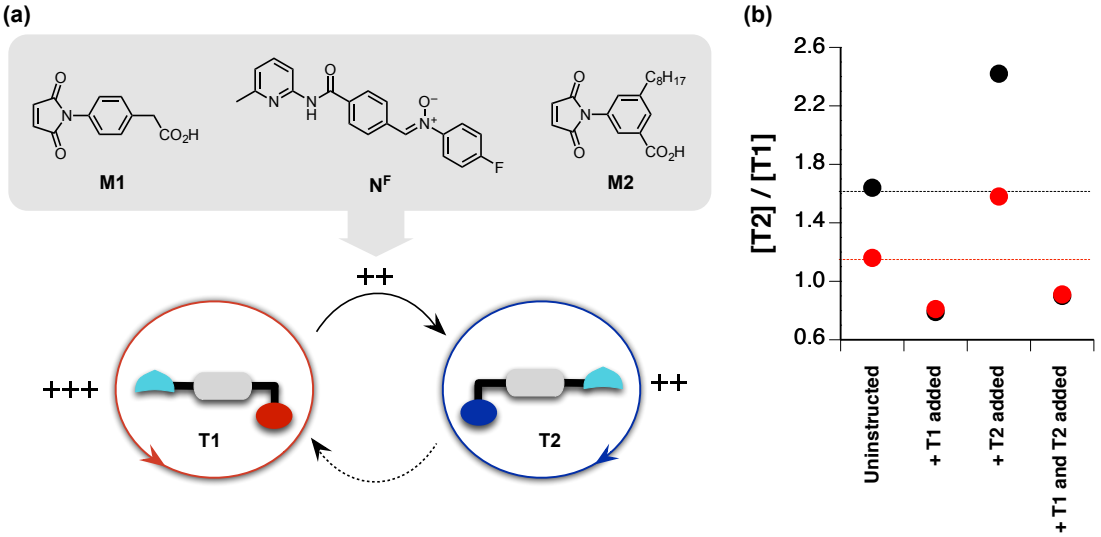


Figure 3.21 **(a)** Three simple chemical building blocks, nirtone $\mathbf{N^F}$ and maleimides **M1** and **M2**, react through 1,3-dipolar cycloaddition reactions to form an interconnected network of two structurally-similar replicators: **T1** and **T2**. A comprehensive set of kinetic experiments examining the formation of both replicators revealed that **T1** is a slightly more efficient replicator than **T2** in isolation. **(b)** Under competition conditions, replicator **T2** was found to be capable of outcompeting **T1** for the shared nitrone component in the uninstructed experiment and that instructed with **T2** preformed template, as evidenced by the ratio of these replciators after both four hours (black circles) and 16 hours (red circles). Dashed lines represent the [*trans*-**T2**]/[*trans*-**T1**] ratio determined in the uninstructed kinetic experiment, for comparison.

Analysis of the kinetic parameters determined through fitting showed that the high replication efficiency of **T1** stems from both a higher template-mediated rate constant and a lower stability of its template duplex, compared to **T2**. Whilst both autocatalytic cycles were shown to be active, only **T1** was found to be able to instruct efficiently the synthesis of **T2** in the crosscatalytic kinetic experiments. The higher stability of the **T2** template duplex relative to **T1** dimer manifested itself also in the inability of this

replicator to catalyse the formation of **T1**—the heteroduplex formed between **T1** and **T2** is more stable than the **T1** template dimer, resulting in a lower proportion of the free, catalytically-active template **T1** available for reaction, relative to the experiment lacking template **T2**.

As a result of the interconnectedness of the components in the network, the individual kinetic behaviour of **T1** and **T2** in isolation is altered when the same replicators are examined under experimental conditions enforcing competition for a limited amount of the shared building block. Kinetic analyses of **T1** and **T2** under such competition conditions showed that while both replicators maintain their ability to self-replicate (**Figure 3.21b**), the replicator that is less efficient in isolation, **T2**, outperforms the otherwise more efficient **T1** in the competition for the nitrone component. The complex interplay of recognition and reaction processes means that the system-level behaviour of the replicating network in a scenario where these replicators interact is challenging to predicted despite possessing the information about the individual replicators in isolation. This outcome demonstrates that in a scenario where both replicators can be formed in parallel, **T2** exhibits a higher fitness—*i.e.* ability to transfer structural information and produce copies of itself—than **T1**. The attempts to alter the outcome of the competition between **T1** and **T2** using preformed templates of these replicators as instruction showed that our ability to increase the selectivity for one replicator over another is limited. The closed nature of the reaction environment (*i.e.* well-stirred batch reactor) employed for the analysis of the two-replicator network does not permit the replication processes to perform efficiently once the building blocks are exhausted—this kinetic barrier prevents either replicator from being able to dominate, resulting instead in the erosion of the initial imbalance between the two replicators. Finally, whilst limited, the selectivity determined for one replicator over another through the analysis of this interconnected **T1**–**T2** network within the closed reaction environment, driven by recognition-mediated covalent capture, can be used as a baseline for comparison to selectivities obtained for the same network under dynamic and reaction-diffusion conditions, which will be examined in the subsequent chapters.

CHAPTER 4

PROBING THE SELECTIVITY IN A TWO-REPLICATOR NETWORK EMBEDDED WITHIN A DYNAMIC COVALENT LIBRARY

4.1 Preamble

The work presented in **Chapter 3** focused on the examination of the reaction network constructed from replicators **T1** and **T2** in a closed environment (*i.e.* within a system following a well-stirred batch reactor model), where these two replicators react in the absence of any additional recognition or reaction processes. Having established the specific relationships between the two replicators comprising the network and the kinetic and recognition parameters underlying them, the work presented in this chapter seeks to expand on these results by exploring the behaviour of the **T1**–**T2** network in an environment where the replicating processes are embedded within a mixture of dynamically-exchanging, interconnected components—the so-called dynamic covalent library (DCL) (**Figure 4.1**).

The notion of dynamic covalent chemistry and its utility in the construction of interconnected complex systems was first introduced, together with several examples, in **Chapter 1**. Within the environment of a dynamic covalent library, the replication and selectivity in the network of competing replicators (**T1** and **T2**) can be studied in a situation where, instead of the preformed nitrone, necessary for the formation of both replicators, only the individual aldehyde and hydroxylamine building blocks required for their formation are present within the library at the time of mixing. Therefore, the 1,3-dipolar cycloaddition reactions of $\mathbf{N^F}$ with the two maleimides, **M1** and **M2**, must drive the formation of the **T1** and **T2** replicators using solely the building blocks distributed amongst the interconverting dynamic pool of components.

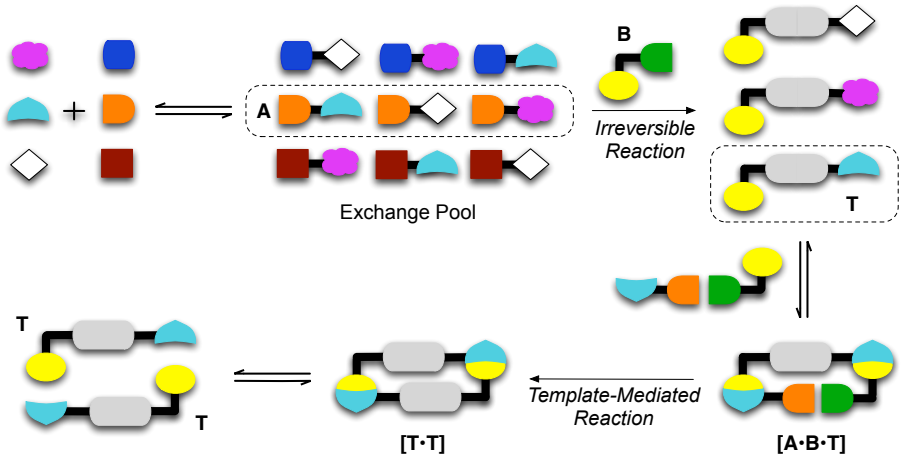


Figure 4.1 Model representing the integration of a dynamically exchanging system with recognition-mediated irreversible reaction processes. In this case, only product **T** is capable of initiating a template-mediated self-replication cycle, thereby removing the component **A**, bearing a suitable recognition site (pale blue) from the exchange pool following its irreversible reaction with component **B**.

4.2 Design of the DCL

The aim of the work presented in this chapter is to develop further the understanding of the behaviour of the two replicators, **T1** and **T2**, referred to as the **T1–T2** network, employing a more complex, dynamic exchange environment of a dynamic covalent library (DCL). The selected environment of a DCL is inspired by the work described[181] previously, and is comprised of four aldehydes, **A** to **D**, and four nucleophiles, **W** to **Z**, affording an exchange pool[a] of 16 condensation products (**AW** → **DZ**), which, together with the unreacted aldehydes and nucleophiles, give a dynamic library of 24 components in total (**Figure 4.2**).

The nucleophile pool is assembled from three amines, **W**, **X** and **Y**, and hydroxylamine **Z**, each tagged with an aryl-fluorine tag, permitting efficient analysis of library samples by $^{19}F\{^{1}H\}$ NMR spectroscopy. Alternative aryl-CF_3 groups were not employed in this library as work described[181,203] previously has shown that the resonances for library components employing such tags could not be resolved successfully as a result of the low sensitivity of their chemical shifts to changes in their electronic environment. This design allows the resonances for the 24 library components, as well as the cycloadducts formed through irreversible 1,3-dipolar cycloaddition reaction with maleimides, to be distinguished by $^{19}F\{^{1}H\}$ NMR spectroscopy (**Figure 4.3**).

[a]Several of the compounds employed in the construction of DCLs have been discussed in the earlier sections of this thesis. For ease of understanding, the numbering of these compounds has been changed in this chapter: aldehydes are labelled from **A** to **D** and nucleophiles **W** to **Z**. Condensation products of these aldehydes and nucleophiles are represented by the combination of their letters. This notation is carried through to the cycloadducts formed by the reaction of nitrones with maleimides.

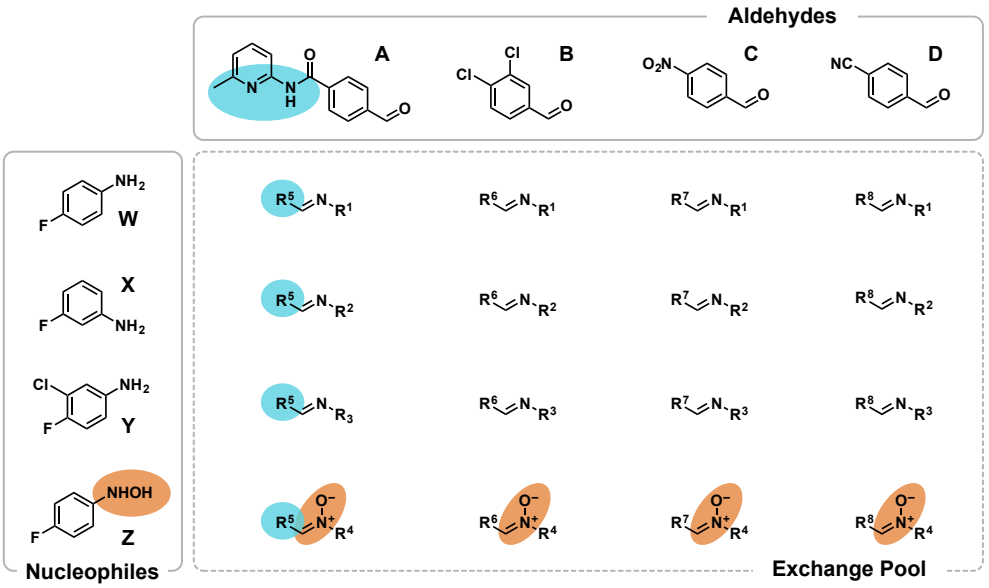


Figure 4.2 Aldehyde and nucleophile components employed in the DCL experiments. Aldehydes **A** to **D** react with anilines **W** to **Y** and hydroxylamine **Z** to produce 12 imines and four nitrone exchange pool components. In the library, only the four nitrones possess the reactive site (orange) necessary for 1, 3-dipolar cycloaddition reactions with maleimides. Similarly, only four exchange pool components formed by reaction with aldehyde **A** bear the amidopyridine recognition site (blue). Only component **AZ**, however, bears both the recognition and reactive element required for template-directed replication.

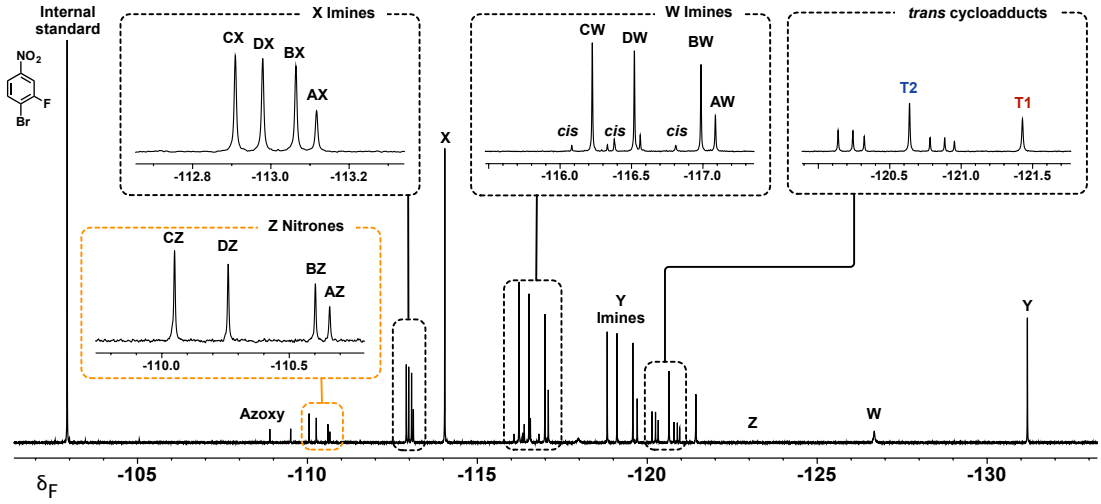


Figure 4.3 Example partial $^{19}F\{^{1}H\}$ NMR spectrum (282.4 MHz, CD_2Cl_2 saturated with *p*TSA monohydrate) of a DCL assembled from four aldehydes and four nucleophiles ([**A**] to [**D**] = [**W**] = [**Z**] = [**M1**] = [**M2**] = 10 mM) and 1-bromo-2-fluoro-4-nitrobenzene as internal standard, after seven days at 5 °C.

The pool of aldehyde and nucleophile components comprising the library was designed to include also the building blocks necessary for the formation of **T1** and **T2** replicators. Specifically, aldehyde **A** was equipped with a 6-methylamidopyridine recognition unit. Condensation of this aldehyde with hydroxylamine **Z** results in the formation of nitrone $\mathbf{N^F}$, referred to as **AZ** throughout this chapter—the key building block for the formation of both replicators. Nitrone **AZ** is the only component in the library possessing both the recognition site and the reactive site required for the recognition-mediated 1,3-dipolar cycloaddition with maleimides **M1** and **M2**. The template-mediated replication of *trans*-**T1** and *trans*-**T2**, explored in **Chapter 3**, was shown to proceed efficiently and diastereoselectively[b], relative to the corresponding recognition-disabled reactions.

When a dynamic pool of components is first assembled from the aldehyde and nucleophile components, there are no condensation products present, and so the library exchange pool can be envisaged as a feeding stock for the formation of nitrone **AZ** as it is removed from the exchange pool over time through irreversible reactions. Presence of **A** and **Z** could enable the recognition nitrone to be produced continuously for a transient period of time—a period limited by the finite amount of library resources and the stability of the different exchange pool components. Overall, the library exchange pool contains four nitrones. In addition to nitrone **AZ**, equipped with a recognition site necessary for self-replication, three nitrones, **BZ**, **CZ**, and **DZ**, capable of reacting only through the much slower and significantly less diastereoselective bimolecular pathways, are present. Ideally, the library can be directed by addition of recognition-enabled **M1** and **M2** maleimides and their corresponding templates, to synthesise the self-replicating templates preferentially, out of the pool of all of the possible cycloaddition products. Selectivity for one recognition-enabled product over another within the system will be examined and compared to the selectivity determined for these replicators within a simple competition scenario, driven by kinetic selection only (**Chapter 3**). Moreover, the ability of the two irreversible, kinetically-controlled processes to influence the dynamic library and shift its composition away from its original equilibrium position, towards preferential formation of these self-replicating products will be analysed.

Prior to examining the feasibility of directing the library to form self-replicating templates preferentially, it was imperative to select conditions that would promote reproducible formation of the exchange pool. Even though previous studies in the Philp laboratory have revealed that $CDCl_3$ works sufficiently well as a solvent for simple

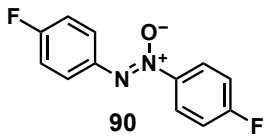
[b]Only the *trans* diastereoisomer is capable of taking part in template-directed replication processes in this system, and, thus, unless emphasis on the *trans* product is required, the notation of the cycloadduct products capable of replication will omit the *trans* notation (for example, *trans*-**T1** will be generally referred to as **T1** throughout the experimental chapters.

kinetic experiments, it does not yield consistently reproducible results under dynamic covalent exchange conditions as a result of the variable water and acid content. Work within the Philp laboratory has shown that CD_2Cl_2, as a less hygroscopic solvent, containing smaller amounts of acidic impurities, is a suitable alternative as a reaction medium. Water and acid were reproducibly introduced to each library by saturating CD_2Cl_2 with *para*-toluenesulfonic acid monohydrate (further referred to as *p*TSA). For details of this procedure, see experimental section in **Chapter 9**. All experiments in this chapter were conducted at the same temperature as employed throughout the kinetic studies in **Chapter 3** (5 °C) in order to permit comparison between results.

4.3 The exchange pool in isolation

As the first step in the analysis of the **T1**–**T2** network within a DCL, the composition of a library exchange pool was examined in the absence of any irreversible reaction processes (*i.e.* no maleimides), permitting determination of the equilibrium position of the DCL. To this end, an equimolar solution of aldehydes and nucleophiles ([**A**] to [**D**] = [**W**] = [**Z**] = 10 mM) was prepared in CD_2Cl_2 saturated with *p*TSA and added to an NMR tube. The sample was allowed to equilibrate at 5 °C and its composition was determined quantitatively by 282.4 MHz $^{19}F\{^1H\}$ NMR spectroscopy after two and seven days. Detailed procedure for the assembly of the exchange pool from the individual building blocks is provided in the experimental section in **Chapter 9**.

For every library component, comparison of the exchange pool composition after two and seven days showed less than 5% difference in concentration, with the exception of the azoxy side-product **90** (**Figure 4.4**), formed by oxidation of hydroxylamine **Z** to the corresponding nitroso compound, and its subsequent reaction with hydroxylamine **Z** itself.



90

Figure 4.4 Structure of azoxy side-product **90** detected in the dynamic covalent library, formed from hydroxylamine **Z** and the nitroso compound produced by *in situ* oxidation of **Z**.

Test studies revealed that while the condensation reactions begin to generate the exchange pool products immediately after mixing, the library reaches its equilibrium position within few hours. Once established, the distribution of the exchange pool is not affected dramatically by a rise in temperature (to RT) in the span of several hours. As described in **Chapter 1**, self-replicating systems have a slow initial rate of product

formation (manifesting as a lag period), but they can perform efficiently from $t = 0$, if instructed with preformed template with suitable catalytic properties. However, in any individual library, there is a large number of components and, thus, the reaction rates are expected to be noticeably slower than in the isolated kinetic experiments (**Chapter 3**). For this reason, DCL analysis was to be undertaken after two and seven days from the time of mixing.

The composition of the exchange pool after two and seven days was shown to be essentially identical by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy, and, therefore, only the distribution typical for an exchange pool after seven days is presented (**Figure 4.5**). Specific concentrations for all exchange pool components, after both two and seven days are available in **Appendix A3**. The equilibrium for the formation of nitrones from hydroxylamine **Z** is located far to the product side, with virtually complete conversion (above 99%) to the condensation products. The high conversion of **Z** is not surprising, as the α-nitrogen lone pair makes hydroxylamine **Z** a very highly reactive nucleophile. Distribution of the hydroxylamine across the formed nitrones reflects the electron-withdrawing ability of the functional groups present on each aldehyde. Nitrone **CZ** is formed at highest concentration ([**CZ**] = 3.3 mM), closely followed by nitrone **DZ** ([**DZ**] = 2.9 mM). Nitrones **AZ** and **BZ** are formed from more electron rich aldehydes and are present at concentration of only 2.3 mM and 2.1 mM, respectively.

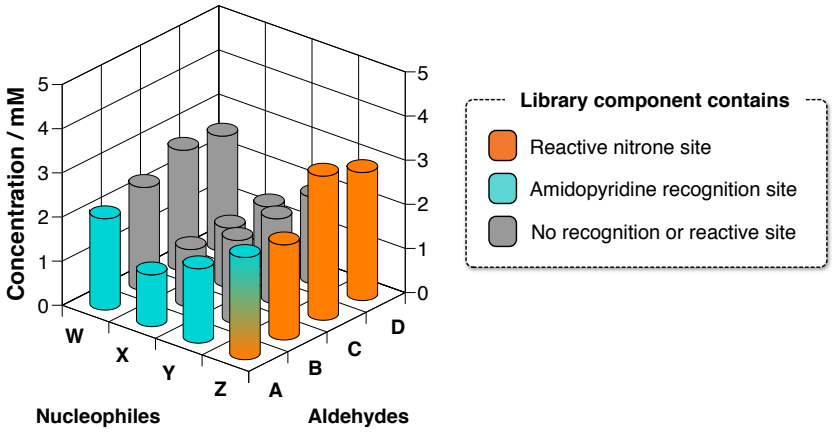


Figure 4.5 Distribution of a dynamic covalent library, comprised from aldehydes **A** to **D** and nucleophiles **W** to **Z**, in the absence of any reactive maleimide components ([**A**] to [**D**] = [**W**] to [**Z**] = 10 mM, in CD_2Cl_2 saturated with *p*TSA) as determined by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy after seven days. Only four library components, nitrones **AZ** to **DZ**, possess the reactive site (orange) necessary for 1,3-dipolar cycloaddition reaction. Similarly, only four exchange pool components formed by reaction with aldehyde **A** bear the amidopyridine recognition site (blue). Only component **AZ** is equipped with both the recognition and reactive element required for template-directed replication. Exchange pool components lacking the reactive and recognition site are represented in grey.

The second best nucleophile in the system is the *p*-fluoroaniline **W**, with overall 88% conversion to imine condensation products. In comparison to **W** and **Z**, the equilibrium

for 4-fluoro-3-chloroaniline **Y** is located further to the reactant side, with only 72% conversion. The least reactive nucleophile is 3-fluoroaniline, **X**, with only 49% of the initial concentration transformed into the **X**-imine exchange pool products.

4.4 DCL coupled to irreversible recognition-mediated processes

With the equilibrium position for this dynamic library established successfully, more complex, instructed library scenarios could be examined. Specifically, in order to perform a comprehensive analysis of the two replicators within the library, a number of experiments need to undertaken (**Figure 4.6**). To begin with, each maleimide, **M1** and **M2**, will be examined alongside a recognition-disabled maleimide **M3** in order to probe the difference between recognition-enabled and disabled reactivity within the library.

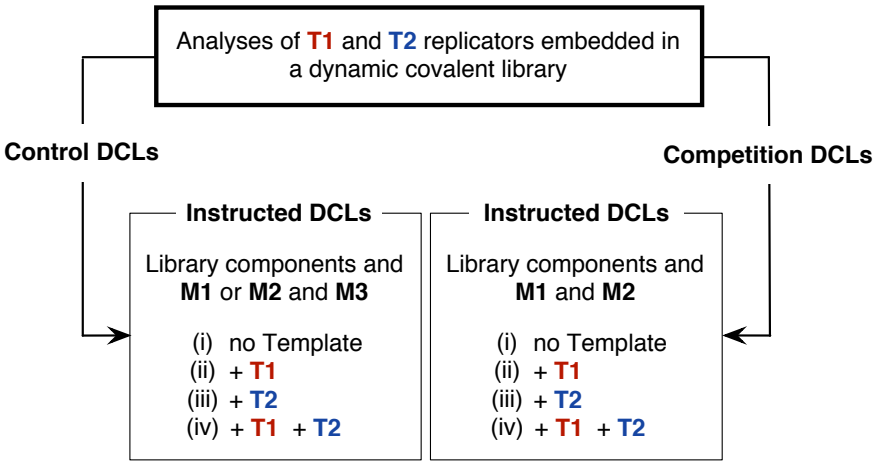
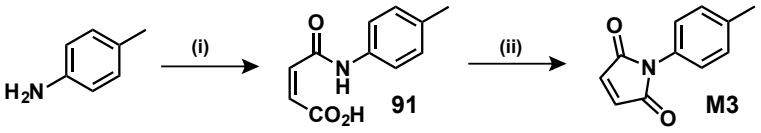


Figure 4.6 A flow chart describing the experiments designed to probe the selectivity in a recognition-mediated self-replicating network embedded in a dynamic covalent library. Formation of each replicator will be examined in the presence of a recognition-disabled maleimide **M3** (with and without instructing template) in a series of control libraries (left). Similarly, the recognition-enabled maleimides **M1** and **M2** will be examined within a competition scenario (right). In each case, the library will be analysed (i) in the absence of template, in the presence of (ii) preformed **T1**, (iii) **T2** and (iv) in the presence of both **T1** and **T2**.

The recognition-disabled control maleimide **M3** was prepared in a two-step reaction from 4-aminotoluene and maleic anhydride. Direct cyclisation of the open intermediate **91** afforded the maleimide **M3** in 52% yield (**Scheme 4.1**).

The behaviour of **M1** and **M2** in the presence of **M3** will be probed in the presence of **T1**, **T2** or both **T1** and **T2** as instructing template. Next, the two maleimides, **M1** and **M2**, will be examined within a competition scenario both in the absence and presence of instructing template (**T1**, **T2**, and **T1** and **T2** simultaneously). Overall, the behaviour of replicators **T1** and **T2** will be probed using 12 different library experiments.



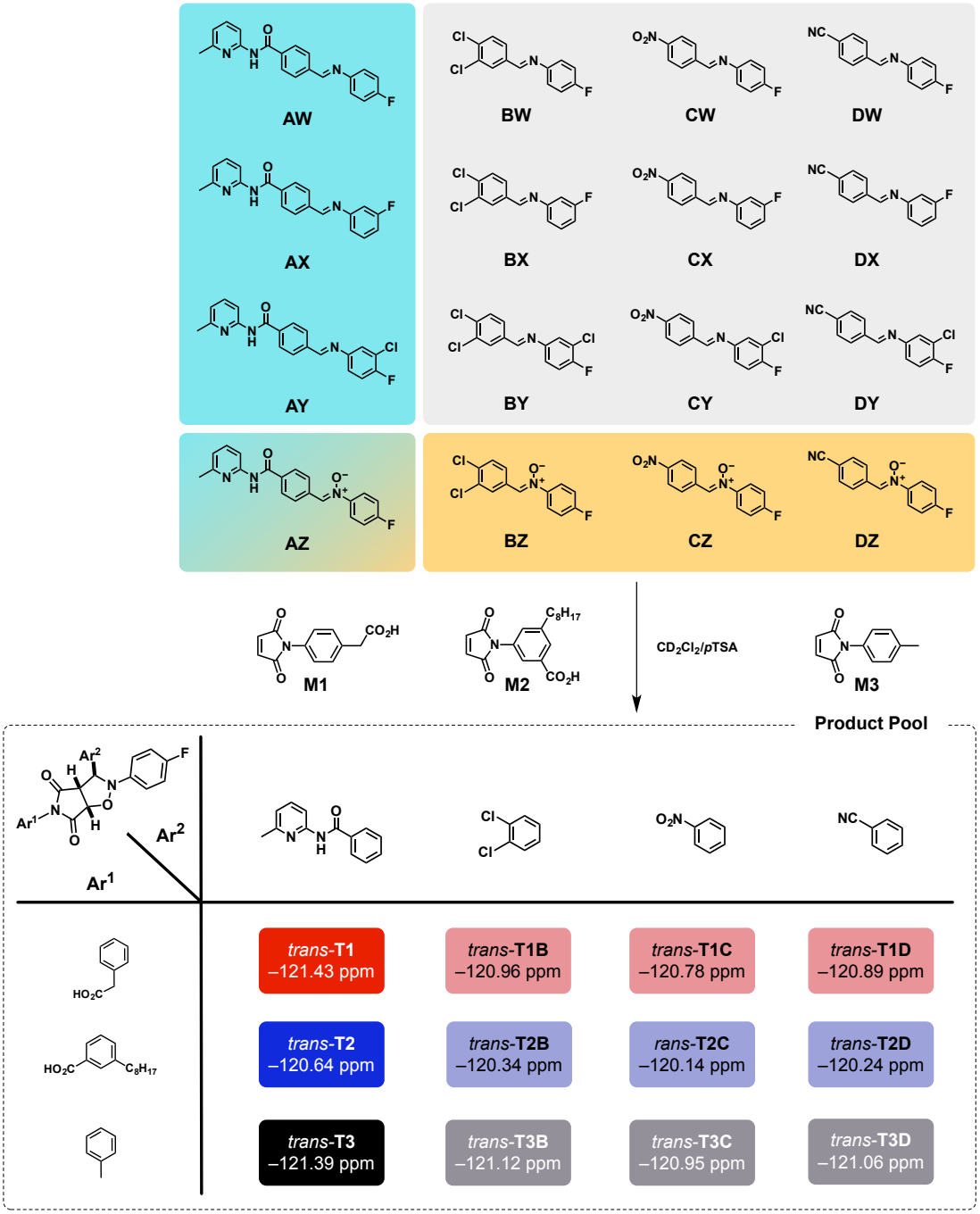
Scheme 4.1 Conditions and reagents employed in the synthesis of maleimide **M3**. (i) Maleic anhydride, AcOH, RT, 3 h, 61%. (ii) NaOAc, Ac_2O, 110 °C, 3 h, 52%.

An overview of the full exchange pool and all 12 possible *trans* cycloadducts formed by the reaction of nitrones **AZ** to **DZ** with maleimides **M1**, **M2** and **M3** is presented in **Scheme 4.2**.

Through preliminary studies, it was possible to establish, that in order to enforce competition, the exchange pool reactive component **Z** has to be present as the limiting reagent. Such reaction conditions, similar to those employed in **Chapter 3** (one equivalent of Nitrone $\mathbf{N^F}$ was used relative to one equivalent of each maleimide) ensure that even if all of hydroxylamine **Z** has reacted or exchanged to form the recognition-enabled nitrone **AZ**, the maleimides will still be present in excess. Therefore, each maleimide was added at a concentration equal to that of the component **Z** (all at 10 mM). Throughout the instructed experiments, 10 mol% of preformed template was employed relative to component **Z**. This amount of template added is slightly lower than the 20 mol% used in template-instructed experiments presented in **Chapter 3**. This change, however, needed to be implemented as the **T1** and **T2** preformed templates, prepared on a preparative scale, were not readily soluble at concentrations above 1 mM in CD_2Cl_2 saturated with *p*TSA.

4.4.1 Control DCLs

In order to demonstrate that the catalytic relationships between the **T1** and **T2** replicators are not affected by the dynamic environment, each recognition-enabled maleimide (**M1** and **M2**) was examined alongside a recognition-disabled maleimide (**M3**) in the absence of template, and using 10 mol% of **T1**, **T2** or both **T1** and **T2** as chemical input. These two primary control libraries (**M1** with **M3** and **M2** with **M3**) allowed the behaviour of the two self-replicators to be examined in the absence of other recognition-mediated processes and the non-recognition mediated selectivity for nitrone **AZ** to be tested.



Scheme 4.2 Structures and $^{19}F\{^{1}H\}$ NMR chemical shifts for all *trans* cycloadduct resonances in the DCL product pool. Reaction of **M1** with nitrone **AZ** produces a self-replicating *trans*-**T1** (red), whereas reaction with **BZ**, **CZ** and **DZ** nitrones results in recognition-disabled *trans* products (pale red). Similarly, reaction with **M2** yields self-replicator *trans*-**T2** (blue) and three further recognition-disabled *trans* products (light blue). **M3** reacts to form *trans*-**T3** (black) and additional recognition-disabled *trans* templates (dark grey). As a result of the ability of *trans*-**T1** and *trans*-**T2** to form recognition-mediated complexes and duplexes, the NMR chemical shifts of the resonances arising from these replicators can vary from those indicated here.

4.4.1.1 M1–M3 library

To begin with, a library containing the standard set of components (aldehydes **A** to **D** and nucleophiles **W** to **Z**, at *ca.* 10 mM, in CD_2Cl_2 saturated with *p*TSA) was instructed with **M1** and **M3** maleimides (no template). After two days (**Figure 4.7a**(i)) *trans*-**T1** and *trans*-**T3** (formed by reaction of nitrone **AZ** with **M3**) were formed at concentration of 1.9 mM and 0.38 mM, respectively. Replicator *trans*-**T1**, as the only recognition-enabled product, dominated the product pool. Comparison of the concentration of *trans*-**T3** formed in the library to the other **M3**-containing cycloadducts (*trans*-**T3B** = 0.30 mM, *trans*-**T3C** = 0.24 mM and *trans*-**T3D** = 0.24 mM for **BZ**, **CZ** and **DZ** *trans*-products, respectively) revealed that the *trans*-**T3** has reached the highest conversion out of all the recognition-disabled products. This subtle difference indicates that the bimolecular rate constant for the formation of *trans* cycloadducts from nitrone **AZ** is higher than the rate constant for the remaining nitrones (**BZ**, **CZ** and **DZ**). The combined concentration of the three **M1**-incorporating recognition-disabled *trans* cycloadducts (*trans*-**T1B**, *trans*-**T1C** and *trans*-**T1D**) reached a combined concentration of 0.85 mM–concentration very similar to that formed by the *trans* products formed upon reaction with **M3** (0.79 mM), indicating that the rate constant for the bimolecular reaction with **M1** is comparable to the respective constant for the reaction with **M3**. While the irreversible reactions have successfully began to transform the library, a significant amount of unreacted **Z** remained distributed across the four nitrones (6.55 mM out of 10 mM).

Comparison of the exchange pool distribution to the distribution determined for a control library after two days (**Figure 4.7a**(ii)) reveals that the DCL is perturbed by the irreversible processes taking place, enforcing library re-equilibration. In particular, the components incorporating the recognition-enabled aldehyde **A** and the reactive hydroxylamine **Z** are removed from the library at the highest rate. The decrease in the concentration of these components is accompanied by an increase in the concentrations of the condensation products formed from aldehydes **B**, **C**, **D** and nucleophiles **W**, **X** and **Y**—imines not requiring **A** and **Z** for their formation.

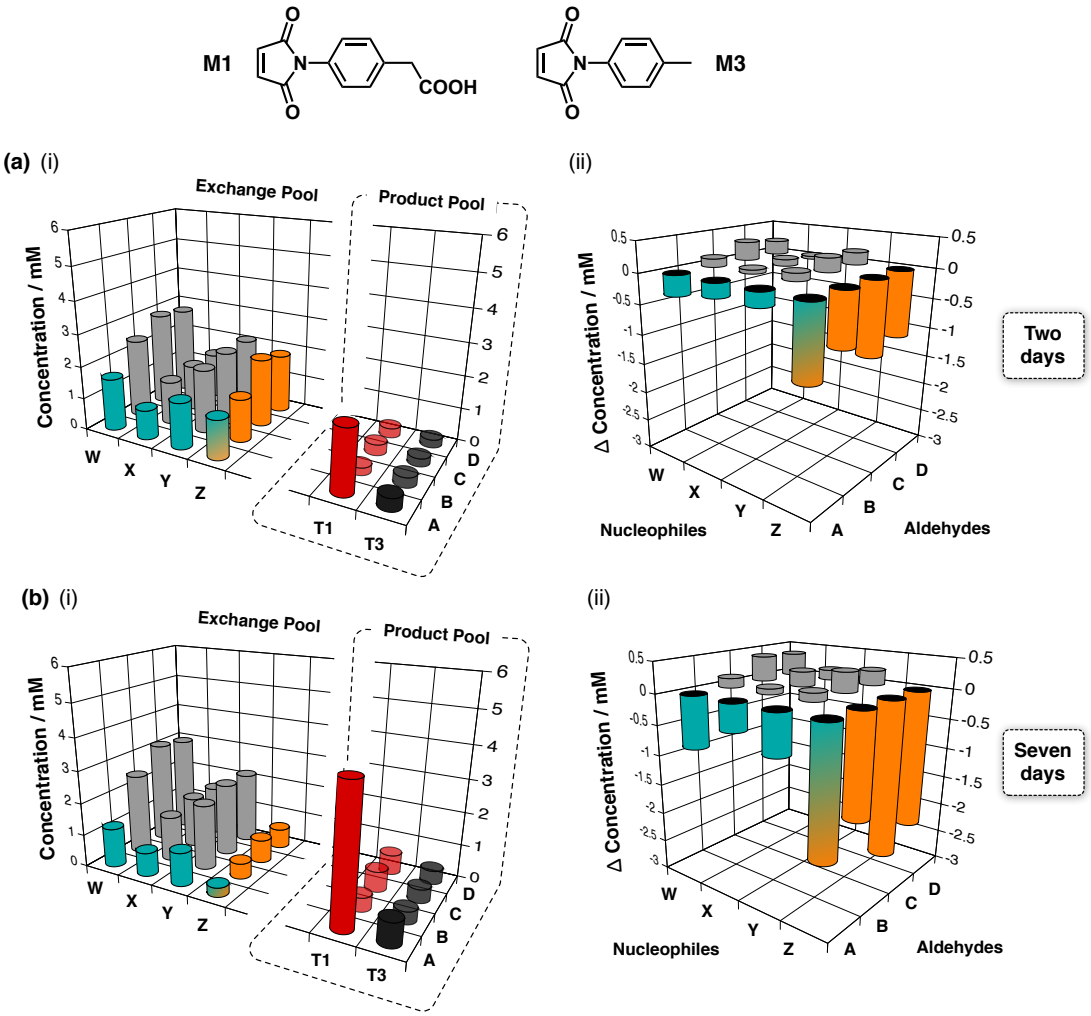


Figure 4.7 A DCL was instructed with a phenylacetic acid maleimide **M1** and a recognition-disabled maleimide **M3**. The graphs show the (i) composition of the exchange pool and the *trans* product pool, where *trans*-**T1** is shown in red and *trans*-**T3** in black, as determined by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy, relative to 1-bromo-2-fluoro-4-nitrobenzene as an internal standard ([**A**] to [**D**] = [**W**] to [**Z**] = [**M1**] = [**M3**] = 10 mM, in CD_2Cl_2 saturated with *p*TSA) after **(a)** two and **(b)** seven days at 5 °C. (ii) Change in concentration of the exchange pool components relative to exchange pool composition determined in the control library (no maleimides) after **(a)** two days and **(b)** seven days. Components labelled pale blue have the amidopyridine recognition site, while orange denotes the reactive nitrone site. Exchange pool components without the reactive or recognition site are marked grey.

After seven days (**Figure 4.7b**(i)), the concentration of *trans*-**T1** and *trans*-**T3** increased to 4.1 mM and 0.72 mM, respectively. Unsurprisingly, the recognition-bearing replicator **T1** remained as the dominant product in the product pool. The *trans*-**T3** cycloadduct continued to be the most abundant **M3** *trans*-product (0.37, 0.41 and 0.41 mM for *trans*-**T3B**, *trans*-**T3C** and *trans*-**T3D**, respectively). Overall, **M1** non-recognition products amounted to 1.64 mM. Coincidentally, the three analogous **M3** products had a combined concentration of 1.64 mM as well. These trends qualitatively agree with the results determined for the same library after two days. Interestingly, the non-recognition products formed *via* the bimolecular pathways increased in concentration less than two-fold—a result consistent with the fact that the rate profile for these reactions is being affected by the progressively decreasing amounts of each nitrone available for the non-recognition reactions. As expected, increase in the proportion of the product pool relative to unreacted material was associated with a decrease in the amount of hydroxylamine **Z**, distributed across the four nitrones and hence still available for reaction, after seven days (2.10 mM).

As noted for the exchange pool distribution determined after two days, the library equilibrium position was clearly perturbed (**Figure 4.7b**(ii)). The concentration of the components encompassing recognition aldehyde **A** and hydroxylamine **Z** in the exchange pool continued to decrease as the irreversible formation of cycloadducts forced the library to re-equilibrate. The overall effect on the library was again reflected in the increased concentrations of the nine exchange pool imines (**Figure 4.7**, grey) formed from aldehydes **B**, **C** and **D** and amines **W**, **X** and **Y**.

4.4.1.2 Template instructed M1–M3 libraries

Having established the effect of embedding **M1** and **M3** maleimides within a dynamic covalent library, the efficiency of the reaction processes associated with these two maleimides and their effect on the DCL composition was further probed by instructing the library with preformed template: + 10 mol% **T1**, + 10 mol% **T2** and both templates (+ 10 mol% each). **Figure 4.8** shows the ratio of [*trans*-**T1**]/[*trans*-**T3**] determined for these four conditions after two and seven days[c]. The ratio of these two products was chosen as a way of assessing the efficiency of the **T1** replicator pathway in the presence of various templates, as it directly reflects the ability of this template to exploit the library components for its own formation, relative to that of the non-recognition products.

[c]Raw data showing the concentrations of each library component across these experiments are supplied in **Appendix A3**.

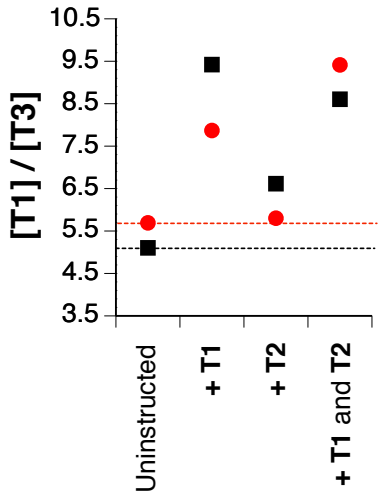


Figure 4.8 Ratio of [*trans*-**T1**]/[*trans*-**T3**] formed in the environment of a DCL across the four examined template conditions (uninstructed, **T1**, **T2** and **T1** and **T2** simultaneously) after two (black squares) and seven days (red circles).

This set of NMR experiments examined the formation of **T1** replicator under instructing conditions similar to those undertaken in **Chapter 3**. These experiments, performed in the absence of other recognition-enabled component **M2**, showed that addition of preformed **T1** results in a marked increase in its formation, particularly at earlier reaction time points (two days). Using **T2** template as an input, however, did not afford the same effect, and, instead, **T2** was shown to be unsuitable as a template for the formation of **T1**. **Figure 4.8** shows that the ratio of [*trans*-**T1**]/[*trans*-**T3**] increases in favour of **T1** replicator most dramatically in the two DCL conditions involving instruction with **T1** template, reflecting the excellent ability of **T1** to template its own synthesis. The template-induced accelerated formation of **T1** affects the ratio of [*trans*-**T1**]/[*trans*-**T3**] more after two days than seven days, most likely because the efficiency of replication slows down as the components are slowly exhausted. The increased efficiency of the **T1** pathway in the presence of preformed **T1** results in a decrease in the availability of hydroxylamine **Z** in other nitrones (**BZ**, **CZ** and **DZ**), slowing down the formation of the recognition-disabled products. The addition of **T2** template exerted virtually no change in the [*trans*-**T1**]/[*trans*-**T3**] after seven days—a result consistent with the kinetic behaviour of **T1** observed previously in the presence of **T2** template. Unexpectedly, the ratio of these products is slightly higher after two days than in the uninstructed library condition. Comparison of the individual concentrations determined for **T1** and **T3** in the uninstructed and **T2** instructed library, however, reveals that the increase does not stem from a higher proportion of **T1** but instated from a decrease in the formation of **T3** product. The lack of change in the concentration of **T1** confirms the absence of catalysis. Nevertheless, it is also possible, that simply the presence of **T2**

as a template equipped with complementary recognition sites, capable of associating with nitrone **AZ** (as well as other **A** containing imines) makes **M1** less available for reaction with maleimide **M3**.

4.4.1.3 M2–M3 library

In the next step, the recognition-enabled benzoic acid maleimide **M2** was examined alongside **M3** in a series of analogous DCL experiments. To begin with, the library sample containing the desired aldehydes and nucleophiles was prepared as described previously and examined after two (**Figure 4.9a**) and seven days (**Figure 4.9b**) in the absence of any instructing template. Analysis of the DCL product pool after two days (**Figure 4.9a**(i)) revealed that the concentration of *trans*-**T2** and *trans*-**T3** reached 1.25 mM and 0.31 mM, respectively. As observed for *trans*-**T1** in the **M1**–**M3** library, *trans*-**T2**, as the only recognition-enabled product with the right conformation for template-directed self-replication, outperformed all the other reaction processes.

The three recognition-disabled *trans* cycloadducts incorporating **M2** were formed at an overall concentration of 1.21 mM, while the corresponding **M3** *trans* products reached 0.74 mM concentration. The overall concentration determined for **M3** products is comparable to the concentration determined for these components in the **M1**–**M3** library. These results suggest that the **M2** bimolecular rate constants for the reactions of **M2** with all four nitrones are higher than the rate constants for **M1** and **M3**, further confirming the information conveyed by the kinetic parameters determined through fitting for **M2** in **Chapter 3**. As observed in the **M1**–**M3** library, the exchange pool still contained a significant amount of unreacted nitrones (combined concentration of 6.1 mM) after two days.

As observed for all the instructed **M1**–**M3** libraries thus far, the irreversible reactions in this system perturbed noticeably the distribution of the exchange pool (**Figure 4.9a**(ii)). The presence of an efficient replicating system in the library was again reflected by a progressive disappearance of components containing aldehyde **A** and hydroxylamine **Z**, and a corresponding increase in the concentrations of the imine products not requiring their formation.

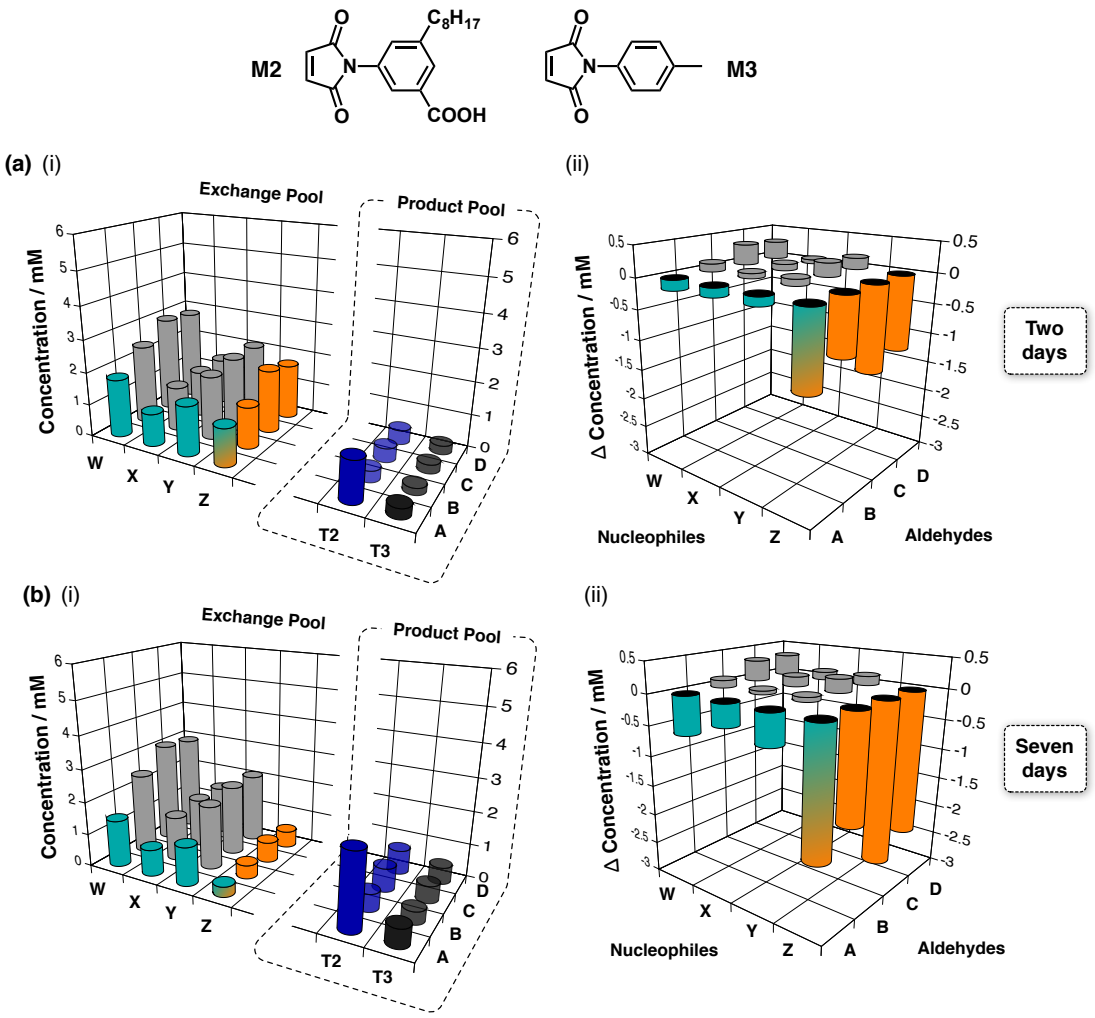


Figure 4.9 A DCL was instructed with a benzoic acid maleimide **M2** and a recognition-disabled maleimide **M3**. The graphs show the (i) composition of the exchange pool and the *trans* product pool, where *trans*-**T2** is shown in blue and *trans*-**T3** in black, as determined by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy, relative to 1-bromo-2-fluoro-4-nitrobenzene as an internal standard (**A**] to [**D**] = [**W**] to [**Z**] = [**M2**] = [**M3**] = 10 mM, in CD_2Cl_2 saturated with *p*TSA) after **(a)** two and **(b)** seven days at 5 °C. (ii) Change in concentration of the exchange pool components relative to exchange pool composition determined in the control library (no maleimides) after **(a)** two days and **(b)** seven days. Components labelled pale blue have the amidopyridine recognition site, while orange denotes the reactive nitrone site. Exchange pool components without the reactive or recognition site are marked grey.

The results determined by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy after seven days (**Figure 4.9b**(i)) showed an increase in the concentrations of *trans*-**T2** and *trans*-**T3** to 2.28 mM and 0.56 mM, respectively. Cycloadduct *trans*-**T2** was the product formed at the highest overall concentration, while *trans*-**T3** remained as the most abundant **M3** *trans* cycloadduct (0.4, 0.5 and 0.5 mM for *trans*-**T3B**, *trans*-**T3C** and *trans*-**T3D** *trans*-**M3** products). Analogously to the results after two days, the **M2** non-recognition products have reached a higher overall concentration (1.97 mM) after seven days compared to the three analogous **M3** cycloadducts (1.40 mM), highlighting the slow formation of **T2** in the absence of template early in the reaction time course. Over seven days, the replication process leading to the formation of **T2** progressed further and the amount of **Z** left distributed amongst the four nitrones decreased to 1.87 mM.

Figure 4.9b(ii) shows the relative change in concentration in the exchange pool after seven days, compared to the exchange pool distribution determined in the library lacking any maleimides after the same time. Yet again, the concentration of imines containing aldehyde **A** as well as the concentration of all four nitrones (formed from hydroxylamine **Z**) decreased dramatically. Concomitantly, the imines not requiring this aldehyde for their formation increased in concentration—a system-level behaviour exhibited by the library as a result of the interconnectedness of the DCL components.

4.4.1.4 Template instructed M2–M3 libraries

With the baseline for the formation of **M2** and **M3** cycloadducts within a dynamic library in the absence of instructional template now established, the same DCL experiments were performed using 10 mol% of **T1**, 10 mol% **T2** and 10 mol% of both templates as input. **Figure 4.10** shows the determined ratio of [*trans*-**T2**]/[*trans*-**T3**] for these four experiments after two and seven days. In accordance with the previous sections, specific concentrations for all library components are provided in **Appendix A3** only.

Examination of the ratio of [*trans*-**T2**]/[*trans*-**T3**] in **Figure 4.10** shows that the trends in the efficiency of the recognition-mediated pathway leading to the formation of *trans*-**T2** replicator relative to the recognition-disabled **M3** counterpart, agree qualitatively with the kinetic data pertaining to this replicator discussed in **Chapter 3**. More specifically, the ratio of [*trans*-**T2**]/[*trans*-**T3**] was found to be the lowest in the DCL condition lacking any instructing template, and increases across all experiments employing template. Both **T1** and **T2** templates can template the synthesis of **T2** efficiently, and, thus, the ratio of [*trans*-**T2**]/[*trans*-**T3**] increases in the presence of either of these templates. The highest ratio of [*trans*-**T2**]/[*trans*-**T3**] was found in the condition employing simultaneous instruction with both **T1** and **T2** template.

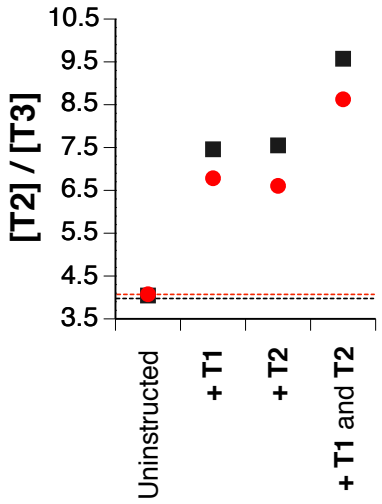


Figure 4.10 Ratio of [*trans*-**T2**]/[*trans*-**T3**] formed in the environment of a DCL across the four examined template conditions (uninstructed, **T1**, **T2** and **T1** and **T2** simultaneously) after two (black squares) and seven days (red circles).

4.4.2 M1–M2 library: T1–T2 competition embedded within a DCL

The two series of control DCL experiments examining the formation of **T1** and **T2** in the presence of non-recognition-mediated maleimide **M3** confirmed that their replication efficiency and catalytic relationships are not affected by the transfer to a dynamic environment. Therefore, as the next step, the formation of **T1** and **T2** replicators was examined within a DCL under competition conditions, in order to compare any change in selectivity relative to that determined for the isolated competition scenario in **Chapter 3**—where the normally less efficient *trans*-**T2** replicator reached a marginally higher concentration than *trans*-**T1** in the uninstructed competition experiment. To this end, a DCL sample was prepared as described previously (no template was added at this stage), and its composition was quantitatively analysed by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy after two (**Figure 4.11a**(i)) and seven days (**Figure 4.11b**(i)).

After two days, *trans*-**T1** and *trans*-**T2** reached 0.86 mM and 1.13 mM concentration, respectively. The resulting overall ratio of [**T2**]/[**T1**] was determined as 1.3—a value more similar to the [**T2**]/[**T1**] ratio (1.2) observed in the closed uninstructed competition experiment (**Chapter 3**) after 16 hours that after 4 hours (1.64). The non-replicating *trans* cycloadducts formed from **M1** were produced at a combined concentration of 0.70 mM whereas the **M2** non-recognition *trans* products were formed more efficiently (1.1 mM)—an overall concentration that is actually higher than the concentration of the self-replicating template **T2**. The higher concentrations of the recognition-disabled *trans*-**T2B**, *trans*-**T2C** and *trans*-**T2D** cycloadducts than the concentrations of the **M1** recognition-disabled counterparts, agree with the higher rate of bimolecular reaction reported for the **T2** replicator in **Chapter 3** as well as in the **M2**–**M3** DCL examined in

this chapter. Despite the similarities in the ratio of self-replicators formed under closed and DCL conditions, the conversion was significantly lower in the dynamic system as a result of the increased number of components. In fact, 5.94 mM of hydroxylamine **Z** remained distributed among the four nitrones after two days (no free **Z** was detected).

In order to determine the effect of the two recognition-enabled maleimides on the distribution of the DCL, the exchange pool composition after two days was compared to the exchange pool equilibrium composition in the absence of the maleimides (**Figure 4.11a**(ii)). As expected, imine components equipped with a recognition site (containing aldehyde **A**) decreased in concentration in the **M1**–**M2** library relative to the exchange pool on its own. As noted previously, this decrease is a result of library re-equilibration taking place, in order to compensate for the decreasing amount of reactive nitrones present in the system—nitrone **AZ** in particular. Specifically, component **A** is gradually being released from its 'storage' in **AW**, **AX** and **AY** imines and is transformed into nitrone **AZ**—the component consumed at the fastest rate. This process results in **W**, **X** and **Y** anilines being released from their respective **A**-imines, attributing for the presence of increased amounts of **B**-, **C**- and **D**-imines containing these components. Because of the irreversible processes in the system, all four reactive nitrones are depleted over time, albeit at different rates.

Examination of the product pool after seven days (**Figure 4.11b**(i)) revealed a further increase in the concentration of *trans*-**T1** and *trans*-**T2** to 1.65 mM and 1.91 mM, respectively. The slight decrease in the resulting [**T2**]/[**T1**] ratio to 1.2 can be rationalised by a gradual decrease in the efficiency of both recognition-mediated replication processes over time as the components required for their formation are depleted, eroding the initial imbalance. Nevertheless, this value remains comparable to the 1.2 ratio observed for purely kinetic selection in the closed system. The total non-recognition *trans* cycloadducts formed from **M1** have reached 1.2 mM. The **M2** recognition-disabled cycloadducts (*trans*) continued to be formed faster, reaching an overall concentration of 1.9 mM. After seven days, only 1.61 mM of **Z**-nitrones remained unreacted in the system. **Figure 4.11b**(ii) represents the relative changes in the concentrations of exchange pool components after seven days relative to the library lacking any maleimides. The magnitude of library perturbation is noticeably higher after seven days than after two days.

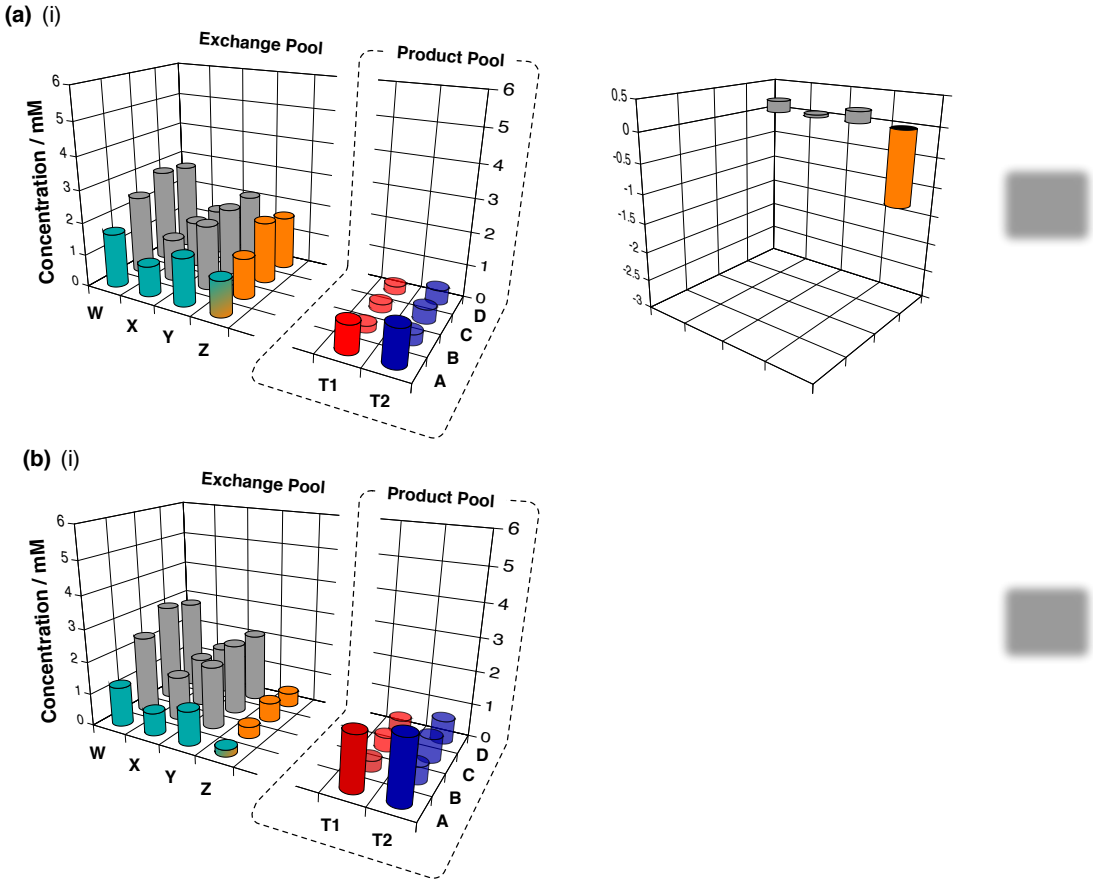


Figure 4.11 A DCL was instructed with two recognition-enabled maleimides: **M1** and **M2**. The graphs show the (i) composition of the exchange pool and the *trans* product pool, where *trans*-**T1** is shown in red and *trans*-**T2** in blue, as determined by 282.4 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy, relative to 1-bromo-2-fluoro-4-nitrobenzene as an internal standard ([**A**] to [**D**] = [**W**] to [**Z**] = [**M2**] = [**M3**] = 10 mM, in CD_2Cl_2 saturated with *p*TSA) after **(a)** two and **(b)** seven days at 5 °C. (ii) Change in concentration of the exchange pool components relative to exchange pool composition determined in the control library (no maleimides) after **(a)** two days and **(b)** seven days. Components labelled pale blue have the amidopyridine recognition site, while orange denotes the reactive nitrone site. Exchange pool components without the reactive or recognition site are marked grey.

4.4.2.1 Template instructed M1–M2 competition libraries

Instructing a dynamic covalent library with **M1** and **M2** maleimides, but not with preformed template, afforded a baseline in product selectivity to which the template directed experiments could be compared. **Figure 4.12** portrays the ratio of [*trans*-**T2**]/[*trans*-**T1**] determined for the four **M1**–**M2** library experimental conditions after two and seven days. As noted previously, the composition of each library exchange pool and product pool is available in **Appendix A3**.

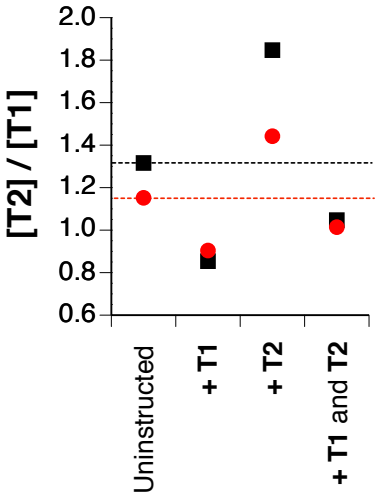


Figure 4.12 Ratio of [*trans*-**T2**]/[*trans*-**T1**] formed in the environment of a DCL across the four examined template conditions (uninstructed, **T1**, **T2** and **T1** and **T2** simultaneously) after two (black squares) and seven days (red circles).

Both *trans*-**T1** and *trans*-**T2** were formed at higher concentrations in the three instructed experiments, relative to the condition with no template input. In the presence of preformed *trans*-**T1**, the ratio of [*trans*-**T2**]/[*trans*-**T1**] decreased as a result of the higher catalytic efficiency of the **T1** formation on the **T1** template compared to the formation of **T2** on **T1**. This ratio in the presence of **T1** was essentially identical after two and seven days. In the presence of *trans*-**T2**, a marked increase in the [*trans*-**T2**]/[*trans*-**T1**] ratio, relative to the template-uninstructed experiment was observed, particularly after two days. The early advantage afforded to **T2** template as a result of the inability of the added **T2** template to crosscatalyse efficiently the formation of **T1** eroded slightly over time. The instructing template exerted the strongest effect on the selectivity at earlier time points, at which the self-replicating reactions generally proceed with low efficiency in the absence of template. Simultaneous addition of both templates simultaneously resulted in a ratio of [*trans*-**T2**]/[*trans*-**T1**] that was very similar to that observed in the **T1**-instructed condition. However, in this case, presence of both template resulted in a more even competition between these two replicators, which were present at a nearly identical concentration after both two and seven days (*ca.*

1). Overall, the results and trends determined for these competition four experiments within a dynamic system agree with the results of the comprehensive kinetic analyses reported in **Chapter 3**.

Having investigated the behaviour of two control libraries, **M1**–**M3** and **M3**–**M3**, and a recognition-enabled competition library **M1**–**M2**, it is instructive to compare the results obtained within these three series of DCL experiments (**Figure 4.13**). Examination of the outcomes from these libraries reveals a stark contrast in the magnitude of product ratios, spanning a range from 3.5 to 10, obtained in the libraries where a single recognition-mediated process performs (**Figure 4.13a** and **Figure 4.13b**) relative to the four **M1**–**M2** libraries where two recognition-mediated reaction processes perform simultaneously (**Figure 4.13c**)—conditions where the product ratios are markedly lower, ranging between 0.8 to 2. **Figure 4.13c** illustrates a dynamic covalent library is instructed with two maleimides, equipped with recognition-sites that enable them to take part in template-mediated reactions, the competition of these maleimides for a shared nitrone **AZ** building block reduces the rate at which the two replicators, **T1** and **T2**, are formed. The addition of preformed template, likewise, results in a significantly larger increase in the ratio of products within the libraries where only a single replicator operates, compared to the libraries with two recognition-enabled maleimides.

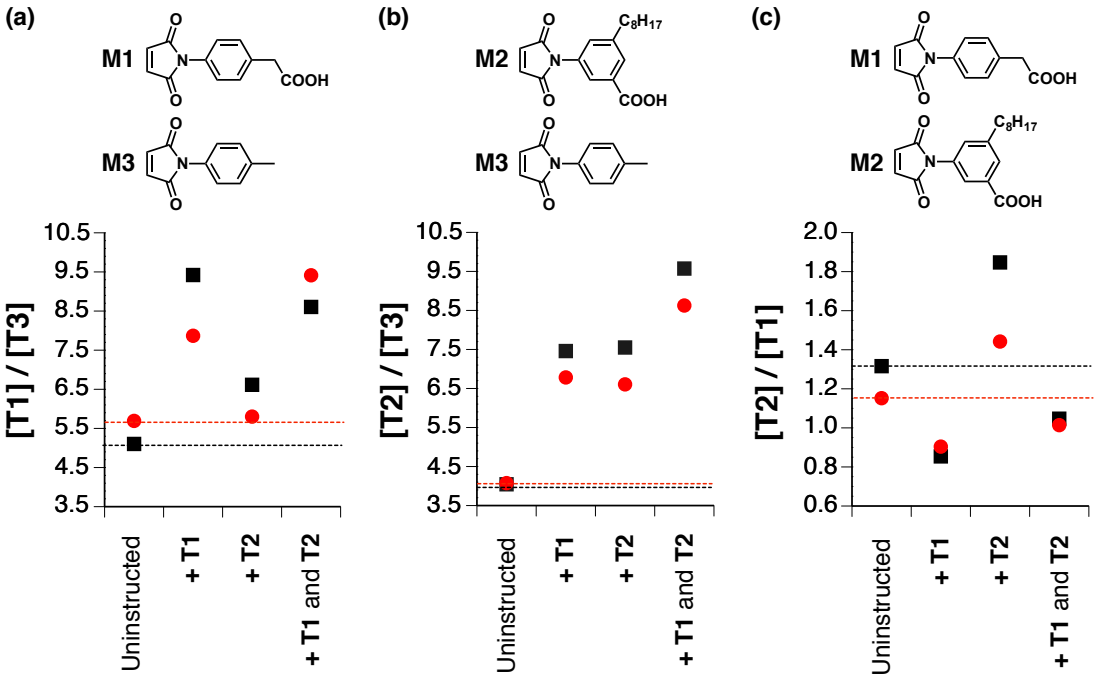


Figure 4.13 Comparison of product ratios determined for **(a) M1**–**M3** libraries, **(b) M2**–**M3** libraries and **(c) M1**–**M2** libraries, across the four examined template conditions (uninstructed, **T1**, **T2** and **T1** and **T2** simultaneously) after two (black squares) and seven days (red circles).

4.5 The limit on selectivity within a DCL

It was envisaged that embedding the self-replicating competing network within a library of dynamically-exchanging components—an environment where covalent and non-covalent selection processes drive network resolution simultaneously—could allow the slight selectivity in the **T1**–**T2** network allowed to propagate. Within the environment of a DCL, both replicators had to manage their formation using the building blocks distributed amongst the entire library—*i.e.* situation where the nitrone required for their formation must be formed through the covalent exchange reaction first. Such conditions, it was hoped, would allow a replicator capable of initiating an autocatalytic cycle at lower template concentrations to consume the formed nitrone faster, enhancing its selectivity. However, the presented results show clearly that combining kinetic selection with dynamic selection did not produce the desired increase in selectivity. In fact, selectivity for a single replicator within a **M1**–**M2** competition library (**Figure 4.12**) under the uninstructed and **T2** template-instructed conditions after two days proved to be lower than the selectivity observed for the analogous competition experiments in the absence of the dynamic pool (**Chapter 3**) after four hours[d] (**Figure 4.14**), governed by kinetic selection. The selectivity in the presence of **T1** and both templates simultaneously was almost identical under both the dynamic and the kinetic conditions. The highest obtained selectivity for a single self-replicator in a DCL was obtained for **T2** in the **T2** instructed scenario, with a final [**T2**]/[**T1**] ratio of 1.85 after two days. The highest selectivity in an environment governed purely by kinetic selection was also achieved in the **T2** instructed experiment—the final [**T2**]/[**T1**] ratio reached 2.38 after four hours. Overall, replicator **T2**, despite being the less efficient replicator in isolation, was found to outperform replicator **T1** in the competition for the shared nitrone in three experimental conditions out of four (uninstructed, **T2** and both templates).

[d]The ratios of [**T2**]/[**T1**] obtained in the DCL after two days were compared to ratios determined within the closed reaction format after four hours as the effect of added template was found to be higher at earlier reaction time points, rather than after seven days and 16 hours.

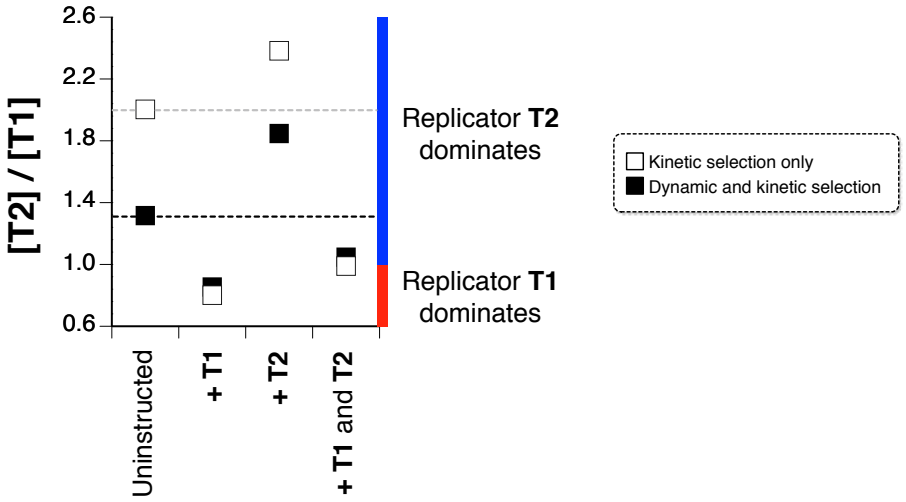


Figure 4.14 Comparison of the ratio of [*trans*-**T2**]/[*trans*-**T1**] formed in the environment of a DCL after two days (black squares) and through kinetic selection only after four hours (white squares) across the four template conditions (uninstructed, **T1**, **T2** and **T1** and **T2** simultaneously) examined experimentally. The red and blue regions highlight the conditions where **T1** and **T2** replicators are the dominant species, respectively. Dashed highlight the ratio of [**T2**]/[**T1**] obtained in both uninstructed conditions.

As mentioned at the beginning of this chapter, the percentage of template used to instruct the reaction mixture had to be reduced from the 20 mol% utilised in the kinetic selection to 10 mol% in the dynamic system. The decrease in concentration of employed instructing template could have contributed to the observed drop in selectivity. Moreover, analysis of the imine components in the exchange pool revealed that while it is possible to push the library away from its equilibrium position, the ability to direct the dynamic system to transform the material available into the recognition nitrone selectively (*via* re-equilibration), remained limited. In order to probe the limits on selectivity within a dynamically-exchanging system, kinetic simulations of the processes within the library, namely the exchange reactions and the recognition mediated reactions were undertaken.

4.5.1 Kinetic simulations

Kinetic fitting of experimental data examining a single recognition-mediated process in isolation, driven either by the formation a binary reactive complex or a template-directed ternary complex can provide access to several key kinetic and thermodynamic parameters underlying its kinetic behaviour. Such kinetic fitting of experimental data pertaining to a replicator, *e.g.* data obtained in the absence of template and in the presence of auto- and crosscatalytic templates was performed for both **T1** and **T2** replicators in **Chapter 3**. The fitting procedure provided information about the bimolecular (k_{bi}) and template-mediated rate constants (k_{uni}) for both the auto- and the crosscatalytic pathways, as well as the duplex association constants (K_a^{Duplex}). Utilisation of these

parameters obtained for **T1** and **T2** replicators in a scenario simulating the uninstructed competition conditions and subsequent comparison of the simulation outcome to the experimental data in **Chapter 3** showed that these two sets of data agree only partially. While the concentration *vs* time profile for the replicator **T2** agreed with the experimental data extremely well, the simulation underestimated noticeably the formation of replicator **T1**.

Simulation of more than a single self-replicating system at the same time presents a more challenging situation than simulation[181] of, for example, two recognition-mediated processes driven by formation of binary reactive complexes. As a result of the complementary nature of the recognition sites employed in the design of self-replicating systems, there is a large number of transient and stable complexes present at any one time within the reaction mixture. Determination of kinetic and thermodynamic parameters through fitting of kinetic data for a single replicator in isolation does not account sufficiently for its behaviour in the presence of another replicating system, and the system-level variation in behaviour is hard to approximate computationally. The kinetic parameters for both replicators cannot be obtained directly by fitting of the competition kinetic data, as the number of parameters fitted would significantly exceed[e] the number of data sets available for the fitting. Nevertheless, despite certain limitations of kinetic fitting, kinetic simulations provide an extremely powerful tool for investigation of the phenomena exhibited by experimental systems—they provide a facile access to conditions that might be extremely challenging to examine experimentally within a particular synthetic framework, not to mention the speed and large number of simulations that can be executed within a short time frame.

Kinetic simulations will be exploited in this chapter with the view to probing the effect of relative catalytic efficiency and template duplex stability on the selectivity achieved for a single product within a network of two competing template-directed reaction processes. In order to approximate the dynamic exchange conditions, a kinetic model was constructed that included the reactions leading to the formation of an exchange pool from aldehydes, **A** to **D**, and nucleophiles, **W** to **Z**. Instead of directly measuring the rate constants for the formation of each exchange pool product, trends in reactivity of the various components were identified and incorporated into the model by means of the k_{off} values for the exchange reactions (**Figure 4.15**). Using the k_{on} and k_{off} values shown in **Figure 4.15**, the simulation of the exchange pool in the absence of any maleimides (**Figure 4.16b**) revealed a good agreement with the experimental data (**Figure 4.16a**).

[e] As noted by Enrico Fermi, John von Neumann has once jokingly remarked[204] "with four parameters I can fit an elephant,[205] and with five I can make him wiggle his trunk".

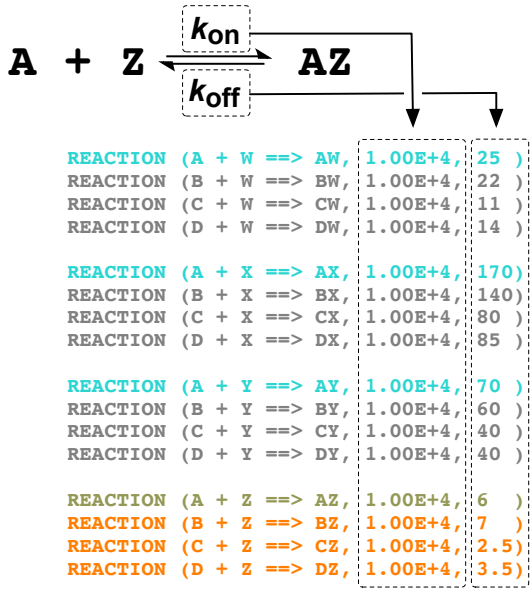


Figure 4.15 Reactions between aldehydes, **A** to **D**, and nucleophiles, **W** to **Z**, leading to the formation of DCL exchange pool. Rate of formation of each component is determined by the ratio of k_{on}/k_{off} rates. The k_{on} rate was set as 1.0×10^4 M^{-1}s^{-1} for each reaction. The k_{off} values (s^{-1}) for the formation of each component were set to reflect the relative reactivity of each component, as determined from experimental DCL exchange pool distribution.

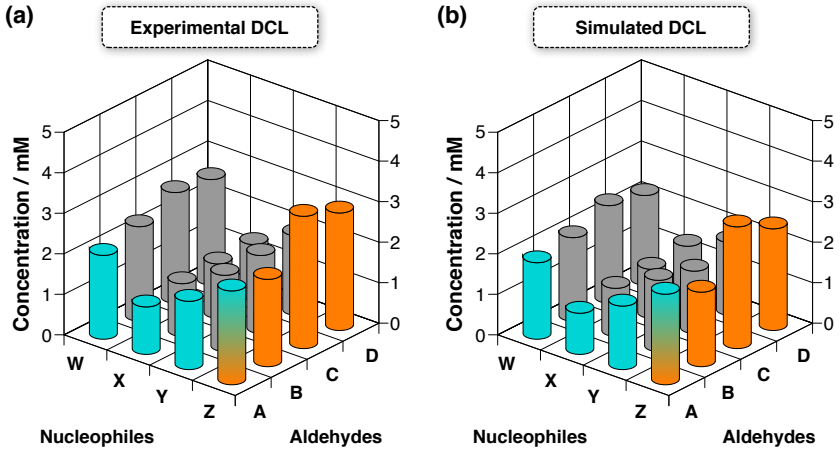


Figure 4.16 Comparison of **(a)** an experimentally determined distribution of a DCL exchange pool after seven days and **(b)** the simulated distribution of the same exchange pool after two days, comprised from aldehydes **A** to **D** and nucleophiles **W** to **Z**, in the absence of any reactive maleimide components ([**A**] to [**D**] = [**W**] to [**Z**] = 10 mM, in CD_2Cl_2 saturated with *p*TSA) as determined by 282.4 MHz $^{19}F\{^1H\}$ NMR spectroscopy.

In the next step, the reactions and interactions pertaining to replicators **T1** and **T2** were incorporated into the simulation. Variation of the rates of exchange reactions with respect to the rates of cycloadditions, *i.e*. set as much faster, intermediate or slow exchange, revealed that the experimental system follows the fast exchange regime (slow and intermediate mode showed significantly worse agreement with the experimental data). The simulated distribution of the dynamic library instructed with **M1** and **M2** maleimides simultaneously (**Figure 4.17**), with the exception of the *trans*-**T1** and *trans*-**T2** recognition-enabled replicators, agreed largely with the data obtained experimentally (**Figure 4.11a**(i)). The concentrations of both **T1** and **T2** replicators produced in this simulation are slightly higher than concentrations produced within the experimental library (after two days). As observed in the simulation experiment examining the **T1** and **T2** replicating network within a competition scenario (**Figure 3.20**) undertaken in **Chapter 3**, the ratio of the *trans* products ([**T2**]/[**T1**]) is more biased towards the **T2** product (1.67), compared to the experimental outcome (1.3). Both simulations employ the same parameters derived through kinetic fitting, and, thus, observation of the same bias is not unexpected. While not ideal, the good approximation of the library exchange pool distribution as well as the general conversion rates show promise for simulations. Nevertheless, as this bias towards replicator **T2** was not observed to the same degree experimentally, the most practical approach to probing the selectivity within a replicating network embedded within a DCL using simulations is to investigate a somewhat simplified analogue of the **T1**–**T2** network, whilst maintaining the network connectivity and the catalytic relationships within it. In particular, because of the complexities associated with simulations of more than a single self-replicating system, the simulations will be directed at probing the effect of varying the ratio of their kinetic effective molarities ($EM_{kinetic}$, introduced as ε in **Chapter 1**) and the thermodynamic effective molarities (EM_{thermo}, **Equation 4.1**) of two replicators, competing for a shared one building block, on the distribution of the product pool and the library as a whole. The parameter EM_{thermo} provides a measure of the enhancement in the stability of the template duplex relative to the stability provided purely by non-cooperative binding between two sets of identical recognition motifs ($K_a{}^{Ind}$). These two ratios of kinetic and thermodynamic EMs will permit the effect of the relative catalytic efficiencies and the relative template stabilities, respectively, to be examined.

$$EM_{thermo} = \frac{K_a{}^{Duplex}}{(K_a{}^{Ind})^2} \qquad (\mathbf{4.1})$$

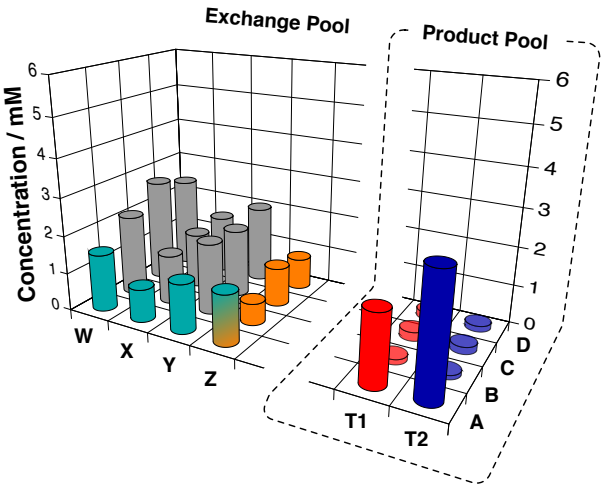


Figure 4.17 Simulated distribution of a DCL instructed with two recognition-enabled maleimides, **M1** and **M2**, determined after two days ([**A**] to [**D**] = [**W**] to [**Z**] = [**M1**] = [**M2**] = 10 mM]). Components labelled blue have the amidopyridine recognition site, while orange denotes the reactive nitrone site. Exchange pool components without the reactive or recognition site are marked grey.

In order to probe the selectivity, a two-replicator (**R1** and **R2**) system analogous to that formed by **T1** and **T2** replicators was designed, governed by the kinetic and thermodynamic parameters shown in **Figure 4.18a**. Maintaining the reactivity and stability trends exhibited by the experimental system, the ratios for various parameters for these two replicators (**R2/R1**) are portrayed in **Figure 4.18b**. The initial parameters chosen for **R1** and **R2** replicators are such that the ratio of their $EM_{kinetic}$ and EM_{thermo} (**R2/R1**) are both 0.1—ratios only slightly lower than those determined for **T2/T1** (*ca.* 0.24). The simulations will therefore examine the effect of varying (i) the ratio of $EM_{kinetic}$ for **R2/R1** (0.1, 1.0 and 10) and (ii) the ratio of EM_{thermo} for **R2/R1** (0.1, 1.0 and 10) (**Figure 4.18c**). Variation of these two parameters will examine how does variation (i) in the efficiency of the autocatalytic pathway and (ii) in the duplex dimerisation constant for a replicator (and the corresponding heteroduplex) translate in to a change in selectivity preference for either **R1** or **R2** replicator. In the nine different simulated conditions (Ia to IIIc), the effect of initial component concentration will be also examined ([C] = 0.1, 0.5, 2.5, 5, 10, 25 and 50 mM). The K_d ($1/K_a$) values for the single point association between the complementary recognition units in **R1** and **R2** replicators are 1.0 mM and 0.33 mM, respectively, and, therefore, will operate progressively less efficiently below these concentrations. The concentration range examined in the kinetic simulations was selected specifically to probe reaction conditions where (i) both recognition-mediated processes operate efficiently ([2.5 to 50 mM] $> K_d$ of the individual binding event of **R1** and **R2**), (ii) one recognition-mediated process operates less efficiently, *i.e.* below its K_d value (K_d of the individual binding event of **R1** $<$ [0.5 mM] $> K_d$ of the individual

binding event of **R2**) and (iii) a concentration where neither of the recognition-mediated processes can operate efficiently (K_d of the individual binding event of **R1** and **R2** $<$ [0.1 mM]). In order to avoid an overly large number of simulations, the parameters pertaining to the crosscatalytic pathways were left unaltered (**Figure 4.18a**) throughout the kinetic simulations. An example script for the kinetic simulations undertaken in this chapter can be found in **Appendix A4**.

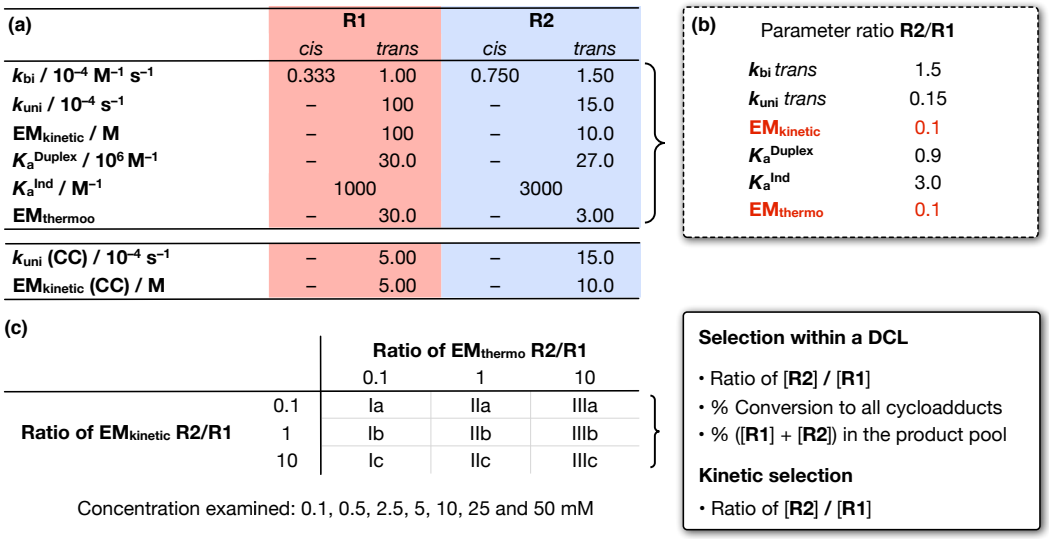


Figure 4.18 **(a)** Kinetic and thermodynamic parameters governing the formation and selectivity within a network of **R1** and **R2** replicators, selected to reflect the catalytic behaviour expressed by the **T1**–**T2** network. CC denotes parameters for crosscatalytic pathways. **(b)** Ratios (**R2**/**R1**) for various kinetic and thermodynamic parameters in the simulated **R1**–**R2** network. Parameter ratios highlighted in red ($EM_{kinetic}$ and EM_{thermo}) will be the two primary parameters varied in the simulations. **(c)** Selectivity in a DCL will be probed by nine different simulation conditions of $EM_{kinetic}$ and EM_{thermo} (Ia to IIIc), with the view to examining the ratio of [**R2**]/[**R1**] found in each system as well as the conversion to cycloadducts. For comparison to selectivity afforded by kinetic selection, the ratio of [**R2**]/[**R1**] will be also examined under conditions lacking the exchangeable pool, commencing from preformed **AZ** nitrone instead. Each of the nine simulation conditions will also examine the influence of initial concentration ([C] = 0.1, 0.5, 2.5, 5, 10, 25 and 50 mM).

The simulations will be carried out with the view to examining the effect of varying these parameters on the ratio of *trans* cycloadducts [**R2**]/[**R1**] formed within a DCL, as well as the overall % conversion to all cycloadducts (formed from the nitrones[f] and the % of the recognition-enabled *trans*-**R1** and *trans*-**R2** products within the entire cycloadduct pool (**Figure 4.18c**). In order to better understand the difference in selectivity afforded

[f]The bimolecular rate constants for the reactions of **M1** and **M2** maleimide analogues with nitrones **BZ**, **CZ** and **DZ** will be set identical to the rate constants for their reaction with nitrone **AZ**. Similarly, the value of the single point association constant for the binding between the two maleimide analogues and all exchange pool components incorporating aldehyde **A** will be confined to the value of K_a^{Ind} for the formation of complex [**AZ**·**M**].

through dynamic selection, the simulated ratio of [**R2**]/[**R1**] within a DCL will be compared to the same ratio determined for conditions lacking the exchangeable pool, utilising a fully formed **AZ** nitrone instead (*i.e.* conditions employed in **Chapter 3**).

Outcome of simulations examining Condition Ia, where the ratio of both $EM_{kinetic}$ and EM_{thermo} is set at 0.1 approximates the most closely the parameters of the DCL system examined experimentally in this chapter. At a concentration of 10 mM, the ratio of *trans* cycloadducts ([**R2**]/[**R1**]) is very close to 1.0. At higher concentrations, the ratio continues to very slowly fall, whilst the opposite trend is observed if the initial concentration is decreased (**Figure 4.19**, Ia). In fact, the highest ratio is observed at 0.5 mM (1.69) instead, revealing a more complex set of rules governing the outcome of the competition than those driving selectivity in a system[181] mediated by formation of reactive binary complexes.

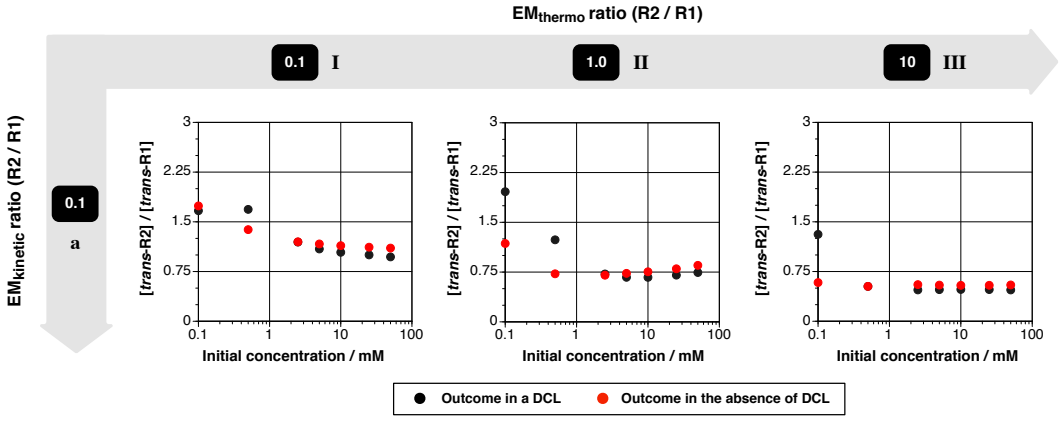


Figure 4.19 Outcome of kinetic simulations probing the influence of initial concentration conditions and the relative template duplex stability (EM_{thermo}) on the ratio of [**R2**]/[**R1**] formed in the product pool in a dynamic system (black circles) and in the absence of dynamically-exchanging components (red circles), after two days. Parameters employed in condition Ia are the most similar to those employed in the **T1**–**T2** system examined experimentally. Simulations were performed using the ISOSIM mode of the SimFit software package. Note that the *x*-axis is presented in logarithmic scale in all cases.

Examination of the same system in the absence of the dynamically-exchanging library revealed a very similar selectivity profile. The most significant difference between the dynamic and purely kinetic selection is that the selectivity in the dynamic system levels off at concentrations < 0.5 mM, whereas in the system driven purely by kinetic selection, the [**R2**]/[**R1**] continues to increase, reaching a maximum value of 1.74 at an initial concentration of 0.1 mM. The dynamic system presents a slightly different scenario that the kinetic selection in that the hydroxylamine component **Z** is distributed amongst four nitrones, as opposed to being fully preformed, and therefore, the effective concentration of **AZ** available for the reactions with the maleimide components is lower than the amount in the kinetically controlled environment. In the dynamic system, the

recognition processes facilitating the formation of **R2** operate more efficiently than those for the **R1** replicator at an initial concentration of 0.5 mM.

In order to explore the effect of relative template duplex stability, the ratio of EM_{thermo} was increased[g] to 1.0 (IIa) and 10 (IIIa) (**Figure 4.19**). Condition IIa presents a situation where **R2** is slightly disadvantaged as a result of increased stability of the duplex [**R2**·**R2**], which reflects in the lower overall selectivity than can be achieved in the concentration range from 0.5 to 50 mM. Nevertheless, the increased stability enabled the system to perform more efficiently at the lowest examined concentration (0.1 mM), resulting in an overall higher ratio of [**R2**]/[**R1**]—1.96.

In condition IIIa, the difference between the relative template duplexes of **R2** and **R1** is further magnified, resulting in a lower selectivity for the **R2** replicator in all cases—a result that can be attributed to the lower relative template stability of the [**R1**·**R1**] duplex (compared to conditions Ia and IIa). In all three conditions, an increase in selectivity for the **R2** species, *i.e.* the product with a higher K_a for the individual binding event but lower overall relative duplex stability (EM_{thermo}) has been observed at the lower concentrations, *ca.* below 0.5 mM—a regime where the low reagent concentration enable replicator **R2** to perform significantly better than replicator **R1** which has a lower ${K_a}^{Ind}$.

The conditions Ia to IIIa explored the effect of the relative template stability by varying the ratio of EM_{thermo} for the two replicators. Next, the simulations were expanded to incorporate the change in the ratio of the $EM_{kinetic}$ for these two replicators, from 0.1 (conditions I) to 1.0 (conditions II) and 10 (conditions III), by altering the recognition-mediated rate constants for the formation of **R2** on **R2** template and **R1** on **R1** template, respectively. Visualisation of the simulation outcome (**Figure 4.20**) from all three $EM_{kinetic}$ conditions on the same *y*-axis scale shows the dramatic increase in selectivity for **R2** replicator in conditions II and III.

The ratio of [**R2**]/[**R1**] within the b and c set of simulated conditions (varying $EM_{kinetic}$) exhibited a similar increase in selectivity (for one replicator over another) at low concentrations. In the dynamic environment, a drop in selectivity at the lowest concentration examined (0.1 mM) was observed in most cases. This drop suggests that again, the recognition-mediated processes for **R2** stop performing efficiently within the dynamic environment. Ultimately, at even lower concentration, the ratio of [**R2**]/[**R1**] would be mostly governed by the bimolecular rate constants for the formation of these two species. In contrast, at high concentrations, the template-directed formation of both

[g]In order to obtain the ratio of EM_{thermo} for **R2/R1** = 1.0, the ${K_a}^{Duplex}$ for [**R2**·**R2**] was increased tenfold. To push the EM_{thermo} ratio further (10), the stability of the [**R1**·**R1**] duplex was decreased by a factor of 10. In each case, the dimerisation constant for the [**R1**·**R2**] heteroduplex was adjusted to account for the changes, using the procedure described in **Chapter 3**.

templates proceeds very efficiently, resulting in a less marked bias for **R2**—the product with the higher K_a^{Duplex} constant.

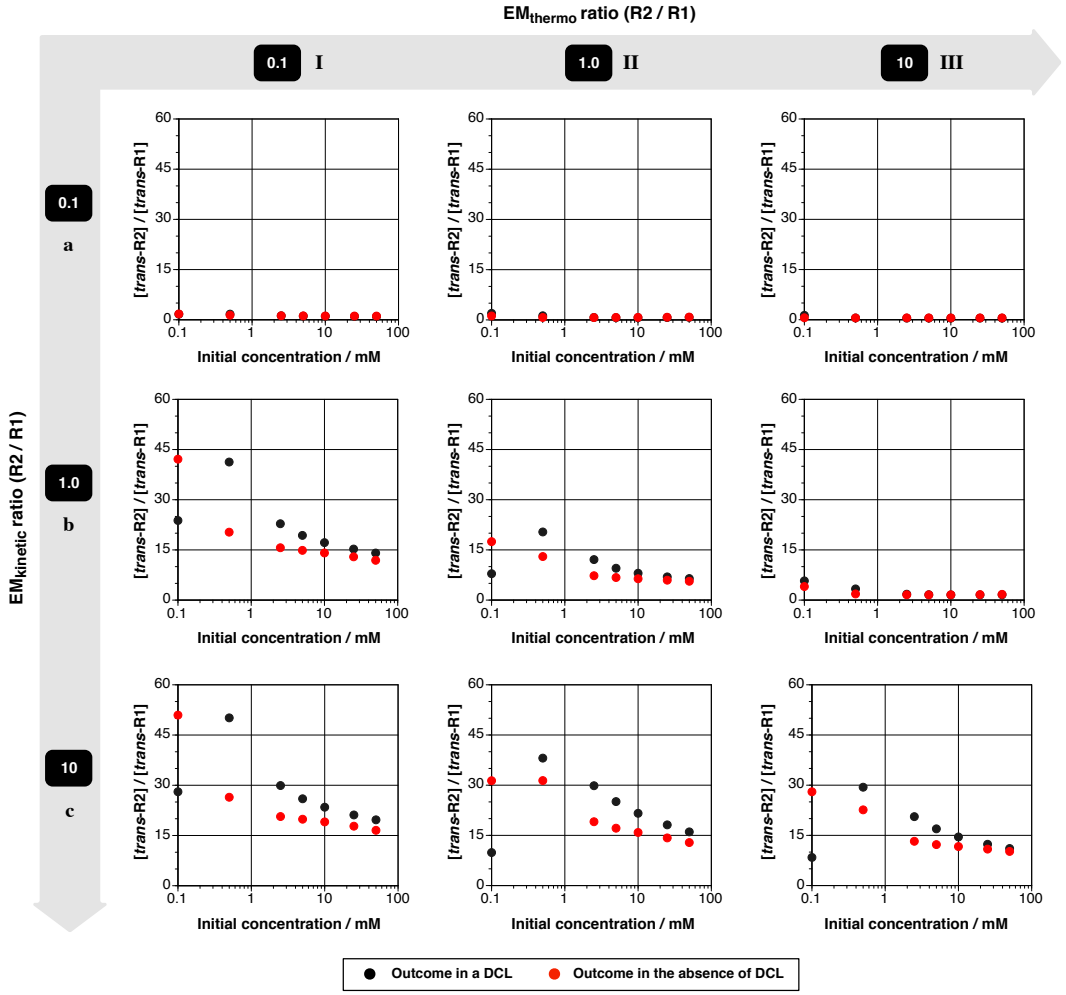


Figure 4.20 Outcome of kinetic simulations probing the influence of initial concentration conditions, the relative template duplex stability (EM_{thermo}, a to c) and the relative template catalytic efficiency ($EM_{kinetic}$, I to III) on the ratio of [**R2**]/[**R1**] formed in the product pool in a dynamic system (black circles) and in the absence of dynamically exchanging components (red circles), after two days. Parameters employed in condition Ia are the most similar to those employed in the **T1**–**T2** system examined experimentally. Simulations were performed using the ISOSIM mode of the SimFit software package. Note that the *x*-axis is presented in logarithmic scale in all cases.

In conditions II and III, where the relative stability of the **R2** duplex increases progressively in comparison to that in condition I, the limit on the ratio of [**R2**]/[**R1**], and thus also the selectivity for the **R2** product, decreases correspondingly. Yet again, such trend shows that replicator efficiency increases with a low duplex stability, which ensures high catalytic turnover.

The DCLs examined experimentally in this chapter showed that the presence of an increased number of building blocks, *i.e.* distribution of the reactive components amongst several species, slows down the efficiency at which material in the exchange

pool is transformed in to the product pool. In each of the nine simulated conditions, the ratio of [**R2**]/[**R1**] was examined after two days—a time frame that would be plausible experimentally, and in fact, also the time after which the DCL examined in this chapter were analysed by NMR spectroscopy. Nevertheless, the concentration conditions examined in these simulations span quite a wide range, and a thus far undiscussed aspect associated with the initial reaction concentration is the conversion, an important factor for consideration if a system is to be examined analytically. To this end, the effect of changing initial concentration as well as the ratios of EM_{thermo} and $EM_{kinetic}$ on the overall conversion to all cycloadducts, and the % of the **R1** and **R2** *trans* recognition-enabled products within the product pool were analysed (**Figure 4.21**).

In an ideal situation, a high ratio of [**R2**]/[**R1**] would be associated concurrently with both high overall conversion and high % of the recognition-mediated products in a system—a combination resulting in a system dominated by replicators **R2** and **R1**, as opposed to unreacted starting materials or recognition-disabled cycloadducts. Examination of these two conversion parameters shown in **Figure 4.21** for each condition reveals a number of trends. The overall conversion to all cycloadduct products after two days increases in the dynamic system with the increasing initial concentration conditions. This increase in conversion is, like the reaction profile of replicating systems, non-linear. The % of the two replicators in the product pool is lowest at the 0.1 mM concentration conditions, *i.e*. situation where recognition-mediated processes leading to the formation of **R1** operate significantly less efficiently than those for **R2**. The difference in efficiency is less marked at 0.5 mM, and the % conversion to **R1** and **R2** increases correspondingly.

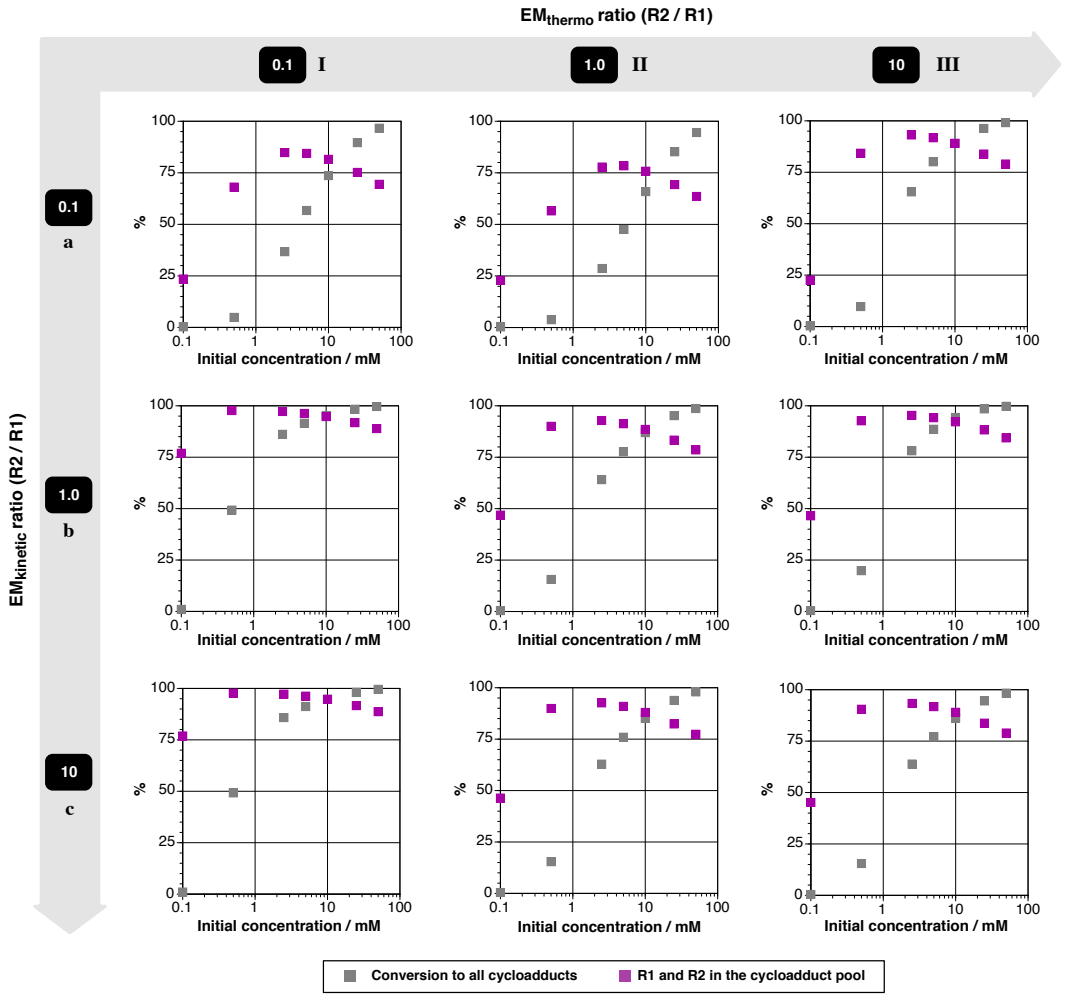


Figure 4.21 Outcome of kinetic simulations probing the influence of initial concentration conditions and the relative template duplex stability (EM_{thermo}) on the conversion to all cycloadducts (grey squares) and the % of the recognition-enabled **R1** and **R2** species in the product pool (purple squares) in a dynamic system, after two days. Simulations were performed using the ISOSIM mode of the SimFit software package. Note that the *x*-axis is presented in logarithmic scale in all cases.

In the majority of the $EM_{kinetic}$ and EM_{thermo} conditions examined, the maximum % of replicators in the product pool is achieved within the 2.5 to 5.0 mM range. Above 10 mM, the % of the recognition-mediated products in the product pool begins to fall—a trend that can be attributed to the higher contribution from the bimolecular reaction pathways. In addition to a lower [**R2**]/[**R1**] ratio, the increase in the relative template stability of **R2** in condition b and c has, as a result of the lower catalytic efficiency, affected also the % of the recognition mediated products. In contrast, the increase in ratio of $EM_{kinetic}$ for the two replicators (conditions II and III) has afforded a higher proportion of the replicators relative to the same concentration examined in condition I.

As suggested earlier, the efficiency at which the material in a dynamic library is transformed into the product pool through irreversible reactions is affected directly

by the distribution of the hydroxylamine within the various nitrone products, **AZ**, **BZ**, **CZ** and **DZ**. A better understanding of the variation of the distribution of the **Z** component in the library exchange pool at different initial concentration conditions can be developed by simulating the formation of the exchange pool from the aldehydes, **A** to **D**, and nucleophiles, **W** to **Z**, in the absence of reactive maleimides. The simulation outcome (**Figure 4.22**) confirms the predicted trend—the concentration of nitrone **AZ** decreases concomitantly with a decrease in the initial reaction concentration. For example, at 10 mM, hydroxylamine **Z** is predominantly present in the form of the nitrone products, where **AZ** accounts for 24% of the **Z**-component. In contrast, only 8.1% of **Z** is transformed into nitrone **AZ** at a concentration of 0.1 mM. A similar trend is observed for nitrones **BZ**, **CZ** and **DZ**, resulting in a significant increase of the free **Z** in the reaction mixture at the lowest simulated concentration ($> 50\%$).

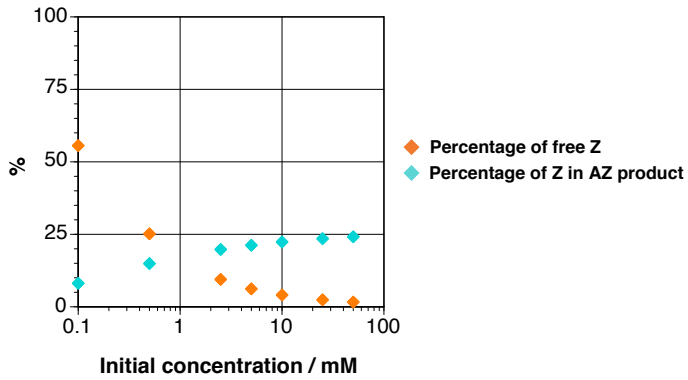


Figure 4.22 Outcome of kinetic simulations probing the influence of initial concentration conditions on % of free unreacted **Z** (orange diamonds) and **AZ** nitrone (light blue diamonds) in the exchange pool. Simulations were performed using the ISOSIM mode of the SimFit software package. Note that the *x*-axis is presented in logarithmic scale in all cases.

The dynamic environment of a DCL presents the replicating network with a noticeably different reaction environment than the reaction format exploited in **Chapter 3**, where the nitrones required for reaction have to be assembled from their building blocks. In an environment lacking the exchangeable pool, nitrone **AZ** is assumed to be infinitely stable, and, therefore, present at exactly the concentration at which it was added originally. In reality, the stability of the added nitrone in its preformed state depends on the amount of water and acid impurities present in the reaction solvent, which means that its actual concentration in the reaction mixture might vary. Nevertheless, as a result of the distribution of **Z** amongst the library components, the amount of nitrone **AZ** available for reaction within the DCL at 10 mM concentration is similar to that available in the kinetic scenario at a concentration of 2.5 mM. This offset in the concentration of nitrone **AZ** in the dynamic and purely kinetic conditions is reflected in the [**R2**]/[**R1**] ratio achieved in the same concentration conditions. As a result of the presence of additional

reactive nitrone components in the DCL, the effect on % conversion is less straightforward. In general however, the conversion at a particular concentration after two days, is always higher in the condition exploiting a preformed nitrone compared to the dynamic environment. A comprehensive overview of the % conversions obtained through kinetic selection is supplied in the appropriate experimental section in **Chapter 9**.

The series of kinetic simulations presented here show that several rules determine the performance of a library as well as the efficiency of the replication processes operating within a DCL (**Figure 4.23**). In condition IIb, the ratio of both EM_{thermo} and $EM_{kinetic}$ for **R2/R1** equals 1.0. The ratio of [**R2**]/[**R1**] formed in the system, is significantly biased towards the **R2** product, as a result of the higher bimolecular reaction rate constant and K_a for the individual binding event. In this case, the highest ratio of [**R2**]/[**R1**], 20.4, is obtained within the dynamic environment, at an initial concentration of 0.5 mM (**Figure 4.23**, grey square).

Radiating from this central simulation are conditions examining the effect of variation in EM_{thermo} and $EM_{kinetic}$. The most substantial increase in selectivity is achieved in condition IIIa, where the $EM_{kinetic}$ ratio is increased to 10, whilst the relative duplex stability (ratio of EM_{thermo}) of **R2** is decreased to 0.1. Clearly, a replicator can perform most efficiently in a competition scenario if it possesses high catalytic efficiency but low template duplex stability. In this particular scenario, a slightly higher selectivity is afforded after two days under kinetic selection conditions. Interestingly, examination of the [**R2**]/[**R1**] ratio over time within the dynamic (0.5 mM) and kinetic environment (0.1 mM) where the highest selectivities were obtained shows that the ratio under these conditions actually peaks after 24 hours and 32 hours, respectively. At these reaction times, the [**R2**]/[**R1**] ratio reaches 54.0 and 53.4 for the dynamic and kinetic selection, respectively, illustrating that the dynamic selection can afford a higher ratio in [**R2**]/[**R1**]. It should be noted, however, that whilst earlier reaction times have the potential to afford a slight increase in the selectivity for a particular replicator, this benefit needs to be weighed against the lower overall conversion obtained at these reaction time points.

The comprehensive set of kinetic simulations presented in this chapter was directed at probing the effect of certain key kinetic and thermodynamic parameters on the selectivity achievable in a system of two replicators embedded within the environment of a DCL, and its comparison to the selectivity achievable through kinetic selection only. The simulations have demonstrated that working at concentrations below the K_d values of the recognition processes can exert a positive influence on the difference in selectivity between two species amplified through template-directed reactions. This positive influence is the most significant at reaction concentrations where the recognition processes for one replicator operate significantly more efficiently than the processes for

another. The most significant selectivity within a competing network of two replicators, **R1** and **R2** (above 50), can be achieved in conditions where the ratio of their EM_{thermo} is as low and the $EM_{kinetic}$ as high as possible—a situation only partially fulfilled by the current **T1**—**T2** experimental system. Ultimately, these simulations reveal the potential selectivity that can be achieved in a network of two competing replicators under various reaction and parameter conditions—selectivity that could be further biased by addition of preformed replicator template as chemical input used to direct the competition towards the formation of a particular product.

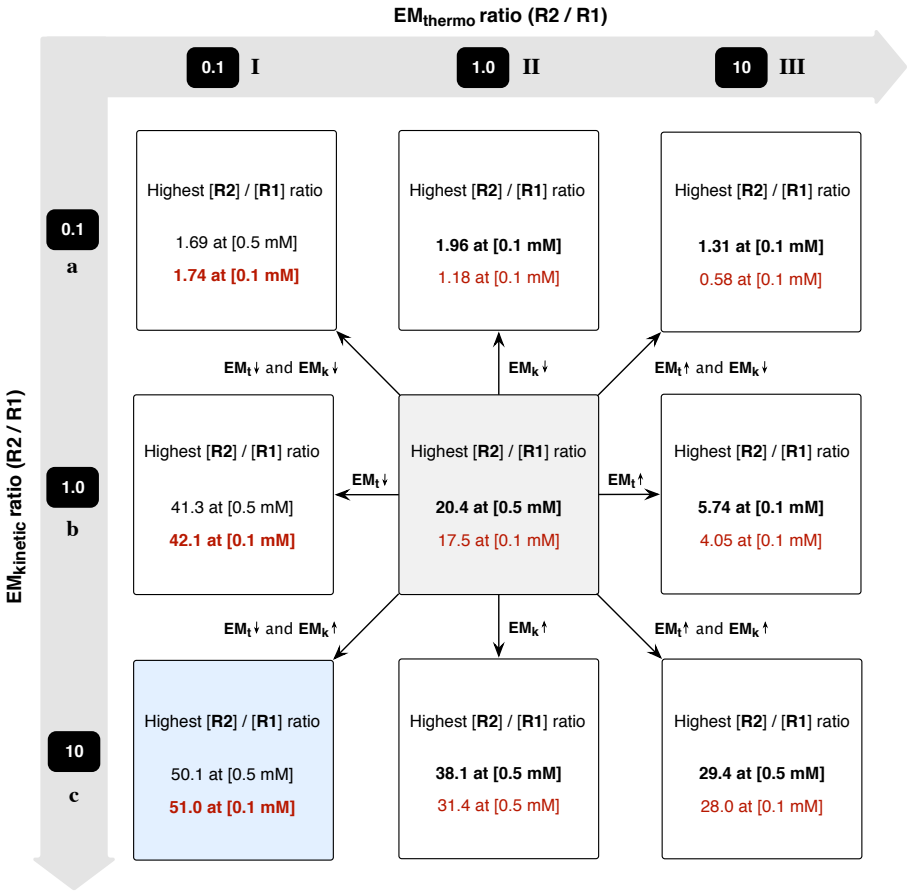


Figure 4.23 Overview of the kinetic simulations probing the influence of initial concentration conditions, the relative template duplex stability (EM_{thermo}) and the relative catalytic efficiency ($EM_{kinetic}$) on the ratio of [**R2**]/[**R1**] formed in the product pool in a dynamic system (black) and in the absence of dynamically exchanging components (red), after two days. In each condition, the dynamic or kinetic environment affording the highest ratio of [**R2**]/[**R1**] is highlighted in bold. The condition resulting in the highest [**R2**]/[**R1**] ratio is highlighted in blue.

In order to investigate how would the addition of preformed template affect the selectivity for one replicator over another, a small set of kinetic simulations was performed. In particular, these simulations examined the effect of adding 10 mol% (calculated relative to the initial concentration of reactants) of preformed (i) **R1** and (ii) **R2** template

on the [**R2**]/[**R1**] ratio in several conditions within both the dynamic environment and that controlled by kinetic selection only. The simulations were performed using the parameters employed in the simulated condition 1C—**R2/R1** ratio of EM_{thermo} = 0.1 and **R2/R1** ratio of $EM_{kinetic}$ = 10—*i.e.* the condition that afforded the highest ratio of [**R2**]/[**R1**] in the absence of any instruction. The template-instructed simulations were performed at two concentrations: 10 mM for both the dynamic and kinetic environment (also the concentration employed in the experiments undertaken in this chapter) and 0.5 mM for the dynamic environment and 0.1 mM for the kinetic condition—two concentrations that produced the highest selectivities in [**R2**]/[**R1**] product ratio within condition 1C.

The outcome of these **R1**- and **R2**-instructed kinetic simulations, as well as the simulation outcome determined for the same conditions in the absence of instruction, is shown in **Figure 4.24**. The results from kinetic simulations undertaken at 10 mM (**Figure 4.24a**) showed that in the presence of **R1**, the ratio of [**R2**]/[**R1**] decreased in both the DCL-containing condition and in its absence. In both the dynamic and kinetic environment, the change in [**R2**]/[**R1**] ratio was markedly less significant in the simulations utilising **R2** as the instructing template—showing a slight increase in the preference for **R2** replicator in each case.

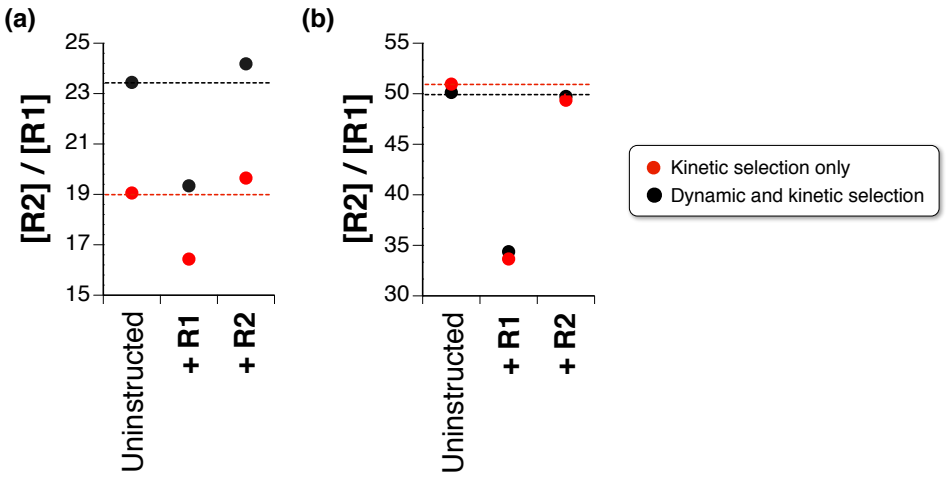


Figure 4.24 Overview of the kinetic simulations probing the influence of added preformed template on the ratio of [**R2**]/[**R1**] formed in the product pool in a dynamic system (black) and in the absence of dynamically exchanging components (red), after two days. Dashed lines indicate the product ratios determined in the absence of template. **(a)** Initial reactant concentration = 10 mM for both the DCL system and the replicators in the absence of DCL. **(b)** Initial reactant concentration = 0.5 mM for the DCL system and 0.1 mM for the kinetic environment. Reaction and association parameters employed in these kinetic simulations are identical to those employed in condition 1C—**R2/R1** ratio of EM_{thermo} = 0.1 and **R2/R1** ratio of $EM_{kinetic}$ = 10. Template was added at 10 mol% relative to the initial reactant concentration employed. The concentrations of products used to determine [**R2**]/[**R1**] were corrected for template added.

The absence of a more significant change in product ratio in these simulations employing **R2** as an instructing template, determined after two days, was probed further by examining the changes in the [**R2**]/[**R1**] ratio as a function of time, within the two simulated DCL systems, at 10 mM (**Figure 4.25a**) and 0.5 mM (**Figure 4.25b**), and comparing them to the changes determined in the uninstructed and **R1**-instructed simulations.

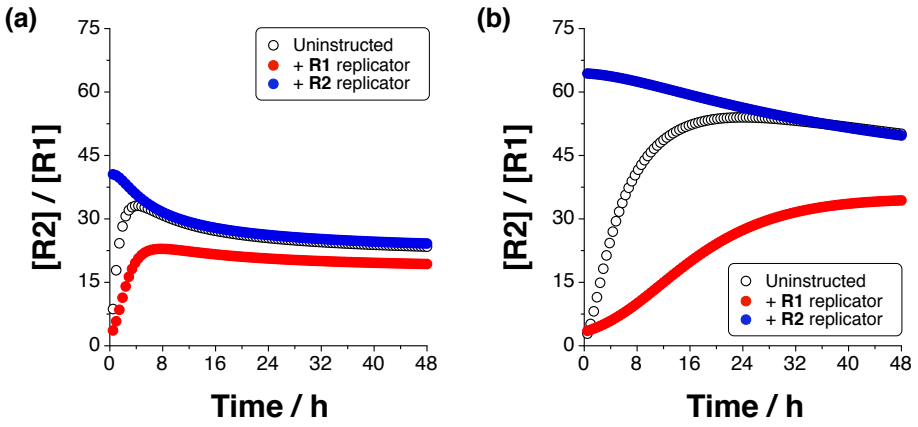


Figure 4.25 Change in [**R2**]/[**R1**] ratio over time in a network of **R1** and **R2** replicators embedded in a DCL at **(a)** 10 mM and **(b)** 0.5 mM. White circles with a black outline represented the [**R2**]/[**R1**] determined in the uninstructed simulations, red and blue circles show the ratios determined in the **R1**- and **R2**-instructed simulations, respectively.

These two plots, showing the [**R2**]/[**R1**] ratio *vs* time revealed that, while the ratio of the two replicators in the uninstructed and **R2**-instructed experiments is similar at the time point examined, two days, the ratio is significantly more different earlier on in the reaction time-course. In fact, the [**R2**]/[**R1**] ratio was the highest at the very beginning of each **R2**-instructed simulation—a time when the presence of the added preformed template plays the most significant catalytic role. Specifically, the preformed template **R2** in these simulations is significantly better at catalysing the formation of itself than the synthesis of replicator **R1**. This difference in catalytic efficiencies means that while **R2** can be formed efficiently on the added preformed template from the beginning, **R1** has to be formed mostly through the template-independent pathways at these early reaction times. **Figure 4.25a** and **Figure 4.25b** show that gradually, the [**R2**]/[**R1**] ratio decreases over time in both **R2**-instructed experiments. In contrast, the simulations lacking any preformed template showed that the lowest selectivity for **R2** replicator is observed at the beginning of the reaction. This product ratio increases over time, reaching a peak at around t = 4.3 hours for $[C]_{initial}$ = 10 mM, and t = 24 hours for $[C]_{initial}$ = 0.5 mM. Interestingly, the [**R2**]/[**R1**] ratios in these **R2**-instructed experiments seem to converge to values very similar to those determined in the uninstructed experiments—this convergence shows the limits on selectivity for one replicator over

another imposed by the closed reaction environment. In the **R1**-instructed experiments, the lowest [**R2**]/[**R1**] ratio (*i.e.* the highest preference of **R1** replicator over **R2**) is again determined at the beginning of each instructed simulation. **Figure 4.25a** and **Figure 4.25b** show, however, that in this case, the [**R2**]/[**R1**] ratios in the **R1**-instructed simulations did not yet converge to the same value as in the uninstructed experiments within the time course examined.

The outcome of kinetic simulations undertaken at 0.5 mM (DCL) and 0.1 mM (no DCL), the [**R2**]/[**R1**] showed that the product ratios across all three template conditions (uninstructed, **R1**- and **R2**-instructed) are noticeably higher than those achieved at $[C]_{initial}$ = 10 mM. Moreover, it is interesting to note that in these simulations performed at low initial concentration, the product ratios determined within the dynamic library and those obtained in the network resolved through recognition-mediated covalent capture processes only are almost identical (**Figure 4.24b**). Again, this observation can be explain by similar changes in the [**R2**]/[**R1**] ratio as demonstrated for the **R2**-instructed experiments at 10 mM. Namely, the highest selectivity for the replicator used as the instructing template is very early on in each reaction, and the selectivity decreases over time, reaching a product ratio similar to that determined in the uninstructed simulations after two days.

These kinetic simulations illustrate that the selectivity for one replicator over another can be biased further by addition of preformed replicator template. Nevertheless, the simulations also showed that the level of selectivity depends significantly on the time point at which the system is examined—in the simulated network of two replicators, the highest selectivity in the presence of preformed template is observed at the very beginning of the reaction. In contrast, the [**R2**]/[**R1**] ratio in the absence of instruction reaches a peak at a time that is dependent on the initial concentration of reagents: the higher the $[C]_{initial}$, the lower the time at which the highest ratio is observed. These simulations demonstrate that the closed reaction environment employed for the examination of the networks of replicators in **Chapter 3** and **Chapter 4** limits the efficiency of the replication processes, which means that the selectivity for one replicator over another does continue increasing over time, and instead, reaches a peak after which this ratio starts to decrease. This limited selectivity, however, means that the the time in the reaction at which the highest selectivity is attained is also generally a time where a significant proportion of the reaction mixture is still present within the form of the reactants (*i.e.* low conversion to products). In contrast, an environment where the efficiency of replication processes would not become self-limiting, the selectivity for one replicator over another might continue on increasing—potentially allowing a system completely selective for a single replicator to emerge.

4.6 Summary

The work in this chapter examined the selectivity in a network of two replicators, **T1** and **T2**, embedded within a dynamic covalent library of imines and nitrones, assembled from four aldehydes (**A** to **D**) and four nucleophiles (**W** to **Z**) (**Figure 4.26a**), both experimentally and through kinetic simulations. The experimental results demonstrated the ability of recognition-mediated self-replicating processes to transform the library exchange pool (**Figure 4.26b**), altering dramatically the distribution of the DCL as a whole. Specifically, in every library instructed with maleimides capable of reacting with the nitrone condensation products *via* irreversible 1,3-dipolar cycloaddition reactions, the exchange pool composition showed signs of library re-equilibration (**Figure 4.26c**), a processes leading to replenishment of the library components incorporating the reactive (**Z**) and recognition (**A**) elements that are being removed from the library at the fastest rate. The presence of components capable of exploiting molecular recognition for their formation affected also the distribution of the product pool—the products equipped with the recognition sites that enabled them to accelerate their formation through template-mediated pathways were produced preferentially at the expense of the recognition-disabled products formed only *via* the slow bimolecular pathways.

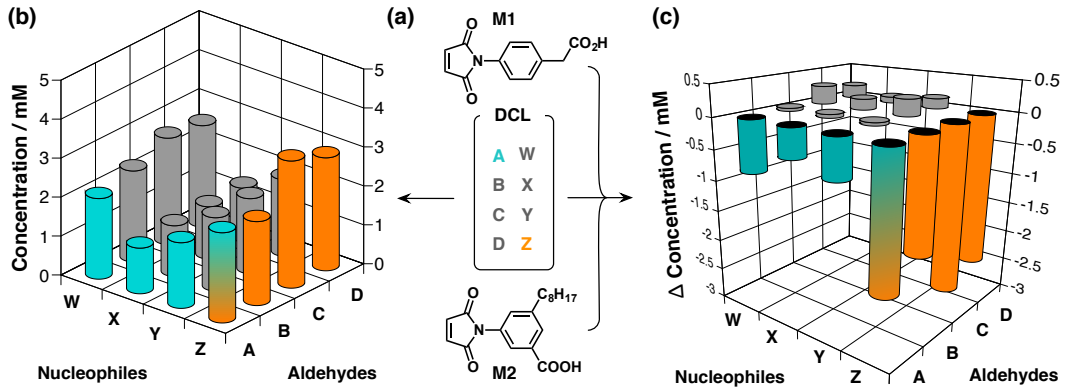


Figure 4.26 **(a)** A dynamic covalent library is assembled from four aldehydes, **A** to **D**, and four nucleophiles, **W** to **Z** at concentration of 10 mM. In CD_2Cl_2 saturated with *p*TSA, the DCL components react reversibly to produce **(b)** an exchange pool of 12 imines and four nitrones, as determined by 282.4 MHz $^{19}F\{^1H\}$ NMR spectroscopy after seven days. **(c)** In the presence of reactive maleimides, **M1** and **M2** (10 mM each), the exchange pool is irreversibly transformed by 1,3-dipolar cycloaddition reactions of the reactive nitrones with the two maleimides, as illustrated by the change in concentration of the exchange pool components relative to composition determined in the absence of maleimides, after seven days. Components labelled blue bear the amidopyridine recognition site, whilst orange denotes the reactive nitrone site. Exchange pool components without the reactive or recognition site are marked grey.

The selectivity in the two-replicator network (assessed by the [**T2**]/[**T1**] ratio) embedded in a dynamic library (**Figure 4.27a**)—*i.e.* a situation where network resolution

is determined by recognition-mediated covalent capture and covalent selection processes simultaneously, proved to be inherently limited and remained governed by kinetic selection under the experimental conditions employed ($[C]_{initial}$ = 10 mM and a ratio of EM_{thermo} and $EM_{kinetic}$ for **T2/T1** around 0.24). Similarly to the kinetic environment, the highest [**T2**]/[**T1**] ratio observed within the environment of a DCL was determined in the presence of **T2** template. Overall, however, the selectivities determined for one replicator over another decreased slightly when the network was embedded within a dynamic environment, compared to those determined through kinetic selection only—an observation that could stem partially also from the decrease in the quantity (mol%) of preformed template used in the DCL instructed experiments, relative to competition kinetic experiments undertaken in **Chapter 3**, as well as a change in the length of time after which the network was examined.

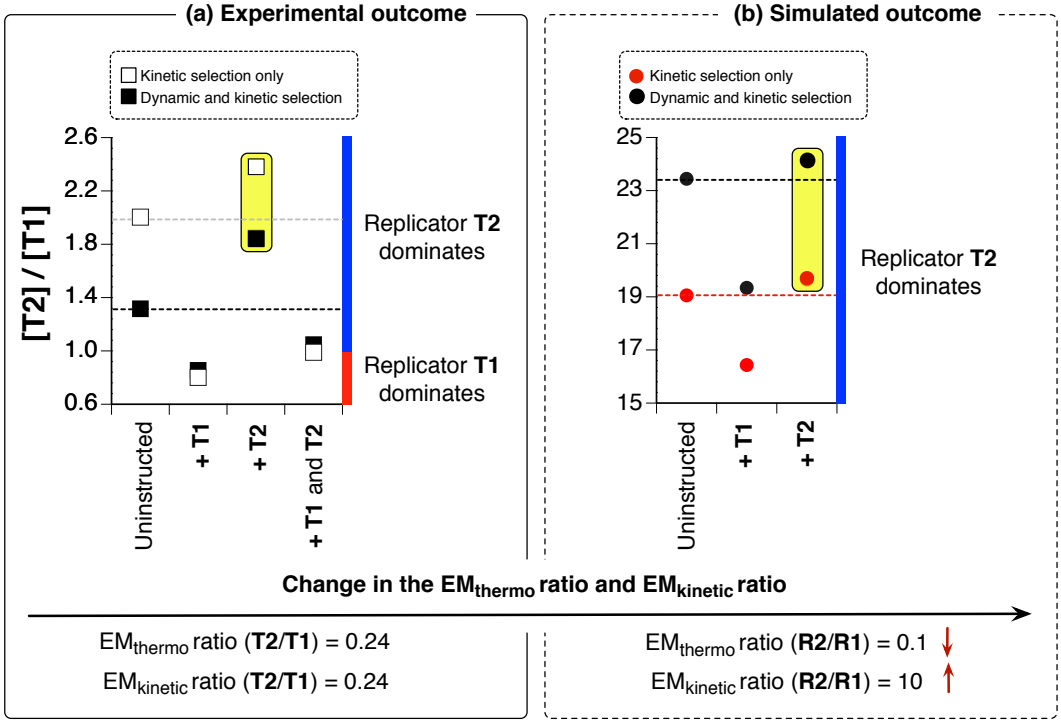


Figure 4.27 **(a)** Ratio of [**T2**]/[**T1**] formed in the environment of a DCL after two days (black squares) and through kinetic selection only after four hours (white squares) across the four template conditions examined experimentally at $[C]_{initial}$ = 10 mM. The ratio (**T2/T1**) of $EM_{kinetic}$ and EM_{thermo} was 0.24 for the replicators examined experimentally. **(b)** Ratio of [**R2**]/[**R1**] formed within a DCL (black circles) and through kinetic selection only (red circles) after two days, in the absence of template, within the simulated condition that resulted in the highest selectivity for replicator **R2**: high ratio of $EM_{kinetic}$ for **R2/R1** (10) and low ratio of EM_{thermo} (0.1). The ratios of [**T2**]/[**T1**] and [**R2**]/[**R1**] achieved in the experimental and simulated network at the same conditions (no template, 10 mM) are highlighted in yellow.

Employing the parameters determined for **T1**–**T2** network as a baseline, the rules governing selectivity for one replicator over another within dynamic and kinetic environment were studied further on a model network of two replicators, **R1** and **R2**

(analogous to **T1** and **T2**), using kinetic simulations. The kinetic simulations revealed that in contrast to reaction networks driven by competing, binary complex-mediated[181] processes, the selectivity in template-driven competition networks embedded within a DCL (relying on the assembly of reaction components in ternary catalytically-active complexes), begins to drop below a critical concentration point—a point where neither of the recognition processes can perform efficiently. Working at concentrations below the K_d for either one or both of the recognition processes can, however, improve markedly the selectivity achieved within a network of two replicators competing for a shared building block.

Outcome of the kinetic simulations confirmed further that the level of selectivity achievable within the examined closed system environment correlate with directly with the difference in the catalytic efficiencies (ratio of $EM_{kinetic}$) and relative duplex stabilities (ratio of EM_{themro}) of the two replicators. A replicator will be able to compete most efficiently for the shared building block if its catalytic efficiency is as high as possible, while the stability of its duplex is as low as possible compared to the ternary complex. Simulation of such network where the difference in the ratio of $EM_{kinetic}$ and EM_{thermo} for the two replicators was significantly larger than in the experimental system (**Figure 4.27a**) revealed *ca.* 10-fold increase in selectivity for one replicator over another (**Figure 4.27b**). A similar enhancement in selectivity, however, was also determined for the same network in the absence of the library, highlighting that the closed reaction environment allows kinetic selection to prevail. The emergence of a replicator capable of propagating the initial imbalance in selectivity that results from the addition of preformed template will require a far-from-equilibrium reaction-environment that will not limit the efficiency of the replication processes operating within. The possibility of employing the environment of reaction-diffusion fronts as a suitable model of open, far-from-equilibrium reaction conditions will be explored in the next chapter.

CHAPTER 5

COUPLING REPLICATION NETWORKS TO REACTION-DIFFUSION PROCESSES

Some of the work presented in this chapter has been published in:

- I. Bottero, J. Huck, T. Kosikova and D. Philp, *J. Am. Chem. Soc.* **2016**, *138*, 6723–6726

5.1 Preamble

The world around us is filled with molecular matter that is constantly interacting and reacting—processes assisted by the motion of molecules through space: random Brownian motion and diffusion. Thus far, **Chapter 3** and **Chapter 4** examined the reaction kinetics of a network of two replicators, **T1** and **T2**, under conditions that follow the well-stirred batch reactor (WSBR) model, *i.e*. a reaction medium where reagents are homogeneously distributed and no exchange of material with the outside environment takes place—the absence of concentration gradient means that diffusion plays little role in this reaction environment. This closed reaction environment, so ubiquitous in the chemical laboratory and so unlike the environment found in nature, places a fundamental limit on the selectivity that can be achieved in a network of two competing recognition-mediated reactive processes and also limits the complexity that can emerge in interconnected systems. In order to move away from the barrier imposed on selectivity by kinetic selection, it is necessary to explore out-of-equilibrium, non-homogeneous reaction conditions that are so prevalent in nature.

Out-of-equilibrium systems[14,206] are much more common than might seem at a first glance—from unpredictable weather cycles to patterns on animal skin, galaxies,

precipitation patterns in rocks, to life itself. For a system to sustain its out-of-equilibrium state, a continuous supply of energy (for example light, heat, chemical energy, *etc.*) and matter is necessary, allowing spontaneous formation of for example self-organising patterns and reaction-diffusion fronts.

Convection cells are a prototypical example of a system that exhibits formation of self-organised patterns arising as a result of an energy gradient. Bénard-Rayleigh convection arises[207] when the space between two planes of glass is filled with a viscous liquid and heat is unevenly applied to the system. In a case where the heat is applied to such a filled plane from the bottom, the fluid close to the source of heat will become increasingly hotter and less dense than the liquid in the top layer, and thus begins to rise from its position. The colder and denser liquid in the upper region, on the other hand, sinks downwards. Combination of these two movements allows convection currents to develop, giving rise[208] to hexagonal beehive-like patterns (**Figure 5.1**). The constant input of heat energy fuels the currents and enables the system to sustain its steady, far-from-equilibrium state.[14] This formation of spatial patterns is observed only once a system is sufficiently far-from-equilibrium, *i.e.* past its instability point. Correspondingly, the state will only last for as long as the external energy is applied and the system has a way of dissipating energy to the surroundings.

Figure 5.1 Example hexagonal patterns formed in Bénard-Rayleigh cell as a result of convection currents driven by a temperature gradient across a layer of oil in between two planes of glass. Figure adapted from Ref. 208.

5.2 Spontaneous formation of self-organised patterns in a reaction-diffusion environment

Formation of hexagonal patterns as a result of convection is an example of a self-organised pattern formed as a result of physical processes. The interplay between reactivity and molecular motion driven by a concentration gradient (*i.e.* diffusion) can allow a plethora of analogous phenomena, such as spontaneous stationary patterns and propagating reaction-diffusion fronts to emerge in chemical[209–211] and biological[212–215] systems alike. As early as 1896, Leisegang observed[216] that mixing of inorganic salts—silver nitrate and potassium dichromate—results in the production of

periodic bands in a layer of gelatinous media. Whilst the striped pattern (**Figure 5.2**) could not be explained by Liesegang at the time, the phenomena of spatial structure formation in precipitation reactions has now been established[217] to arise as a result of the interplay between reaction, diffusion and precipitation. In this particular example, concentric rings are produced when salts react to form insoluble silver dichromate.

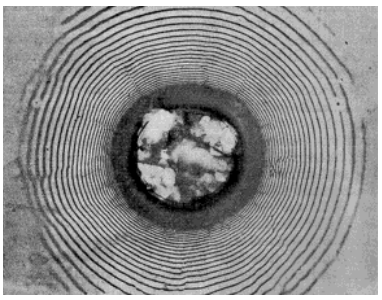


Figure 5.2 Typical Liesegang rings formed as a result of addition of a small drop of silver nitrate to a gel media containing potassium dichromate. Interplay of reaction and diffusion gives rise to periodic striped bands of insoluble silver dichromate. Figure reprinted with permission from Ref. 217. Copyright 2003 American Chemical Society.

In 1952, Turing proposed[218] a mathematical reaction-diffusion model, explaining the spontaneous evolution of spatially heterogeneous patterns in systems comprised of components that are both reacting and diffusing. Turing showed that a theoretical chemical system, formed by an activator and an inhibitor, though initially exhibiting no spatiotemporal phenomena, can evolve towards instability and ultimately, towards formation of spatial patterns. In the system described by Turing, the emergence of patterns was dependent on the diffusion of the inhibitor being greater than that of the activator. Turing suggested that in a scenario where such a reaction-diffusion system was built of the so-called morphogens[a], cells could recognise the pre-pattern of morphogens formed and would respond in a manner that leads to formation of spatial structure within a tissue. Turing reasoned that this process underlies the chemical basis of morphogenesis and is the mechanism behind many of the patterns observed in the animal kingdom. At the time of its conception, Turing's theory was received with scepticism—there was little evidence for the existence of morphogenic compounds. Nevertheless, since then, theoretical simulation studies employing Turing's reaction-diffusion model have been shown[219] to replicate the patterns observed in nature extremely well (**Figure 5.3**).

One of the oldest and the most controversial examples of spatiotemporal phenomena reported is known as the Belousov-Zhabotinsky (BZ) oscillating reaction. In 1958, Belousov, a Soviet chemist, was investigating a reaction mixture containing cerium as a catalyst for the oxidation of citric acid by bromate that was meant to mimic

[a]Morphogen is a signalling molecule that diffuses from the tissue of its origin within an organism and its concentration gradient affects cells exposed to it.

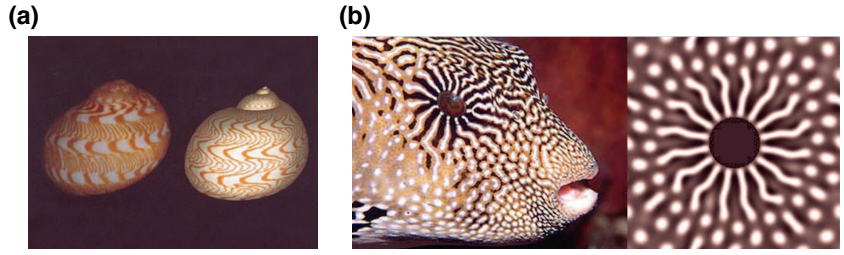


Figure 5.3 Turing pattern observed in a **(a)** seashell and in **(b)** a fish. Patterns shown on the left-hand side are those observed in real organisms, whereas images on the right-hand side illustrate patterns generated through a simulation of a reaction-diffusion system. Figure adapted from Ref. 219. Reprinted with permission from AAAS.

features of glycolysis. Intrestingly, Belousov found[220] that the mixture of reactants he was investigating kept periodically changing colour from clear to yellow. The results reported by Belousov were widely rejected by his contemporaries, on the grounds that the findings contradict the second law of thermodynamics—a law stating that the total entropy in the universe, or in simplified terms the total disorder, must increase over time. Belousov's discovery gained acceptance only after Zhabotinsky demonstrated[221,222] the oscillating nature of the reaction by employing malonic acid instead of citric acid, which improved the visualisation dramatically—changing periodically between red and blue (**Figure 5.4**). A crucial development in the understanding of the non-linear dynamic behaviour underlying the BZ system came with the proposal[223,224] of a mechanism for the BZ reaction by Field, Körös and Noyes.

This is a placeholder for the following image. The image cannot be displayed here due to copyright restrictions.

doi: 10.1387/ijdb.072484vv

V. K. Vanag, I. R. Epstein, *Int. J. Dev. Biol.* **2009**, *53*, 673–681. Fig 1.

Figure 5.4 A series of images illustrating the spatiotemporal patterns exhibited by a Belousov-Zhabotinsky oscillating system over time. Figure taken from Ref. 210

Numerous examples[209,225–229] of oscillations and wave phenomena have been demonstrated in inorganic systems since the first report of the BZ system. Reports of organic systems, exhibiting such behaviour, however, are much more scarce—the literature reveals a single example of a propagating reaction-diffusion front based on small organic molecule, observed[230] in the autoxidation of benzaldehyde. The lack of examples exploiting autocatalysis based on organic systems is in stark contrast to the numerous examples of spatiotemporal dynamic phenomena, ranging from the black

and white stripes on a zebra, to BZ-like waves formed[231–233] by the *Dyctiostelium discoideum* (**Figure 5.5**), present in nature.

This is a placeholder for the following image. The image cannot be displayed here due to copyright restrictions.

J. J. Tyson, J. D. Murray, *Development* **1989**, *106*, 421–426. Fig 1.

Figure 5.5 Oscillating waves formed by aggregating cells of *Dyctiostelium discoideum* in a starvation state as a result of low cAMP concentration. Figure adapted from Ref. 231

Cells of this particular species of slime mould form spiral patterns in a response to starvation-induced chemical wave of cyclic adenosine 3,5-monophosphate (cAMP) propagating through a *Dyctiostelium discoideum* colony. The oscillating pattern stems from cAMP binding to the cell receptors, resulting in their transient desensitisation, producing a travelling wave that is strictly one-directional in character.

More recently, several examples of reaction-diffusion fronts exploiting RNA[234,235] and DNA[236,237] have been reported. In 1989, McCaskill and co-workers reported[234,235] the first example of a reaction-diffusion front observed in a system based on replicating ribonucleic acid (RNA). The propagating RNA fronts were initiated by addition of preformed molecules of RNA at a particular location in an essentially two-dimensional capillary tube reactor, containing a solution of RNA polymerase, nucleotide building blocks and buffer (**Figure 5.6**). Interestingly, McCaskill and co-workers were able to demonstrate that the fronts can emerge stochastically and are capable of evolving as they progress in space over time.

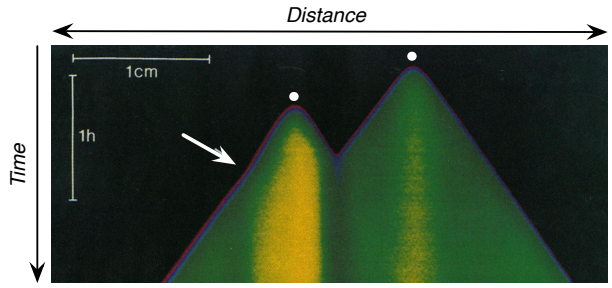


Figure 5.6 A plot showing the distance travelled over time by a spontaneous propagating RNA wave front initiated by addition of two single RNA molecules (white dots). The white arrow indicates a change in the velocity of the RNA wave. Concentration of RNA is represented using a colour scale, where black denotes the lowest and orange the highest concentration. Figure adapted with permission from Ref. 235. Copyright 1993 National Academy of Sciences, USA.

Until 2013, this study was the only example of its kind. Rondelez and co-workers reported[236] an example of travelling concentration waves (**Figure 5.7a**) in a biochemical

network exhibiting predator-prey type of oscillations. The network presented in this study is an extension of their previous work[238] on DNA-based predator-prey systems, employing carefully designed DNA oligonucleotide-based molecules (**Figure 5.7c**), connected by a shared encoding sequence. Namely, the predator-prey network was constructed from three components: prey **N**, predator **P**, and grass **G**—the template required for growth of the prey (**Figure 5.7b** and **Figure 5.7c**).

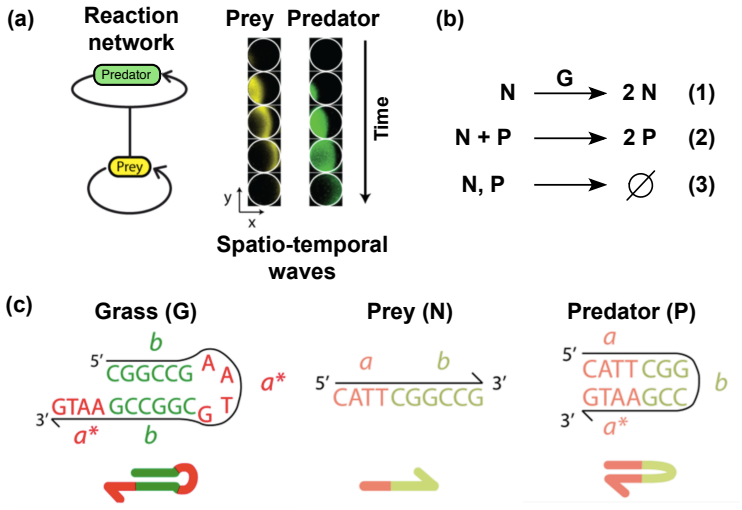


Figure 5.7 **(a)** A biochemical network exhibiting predator-prey type of oscillations, connected by a common DNA oligonucleotide sequence capable of forming a propagating reaction-diffusion front. **(b)** Three reactions at the core of the predator-prey network: (1) autocatalytic growth of prey on the grass template, (2) autocatalytic growth of predator, with a consumption of prey, and (3) decay of predator/prey. **(c)** Structure of grass (**G**), prey (**N**), and predator (**P**). Complementary DNA sequences are highlighted in the same colour, whereas dark and light shade represent regions that can and cannot be destroyed through the action of an exonuclease. Figure adapted with permission from Ref. 236. Copyright 2013 American Chemical Society.

The reactivity within this network is controlled by three purified enzymes: a polymerase[b], a nicking enzyme[c] and an exonuclease[d]—in the absence of these three enzymes, the replication reactions in the network would stall. The prey utilises the grass for its formation, and in turn, the predator consumes the prey component in order to form another molecule of itself. Both predator and prey can decay through the action of the exonuclease enzyme. By examining the molecular network in reaction-diffusion media—within the environment of an unstirred, 8 mm wide and 200 µm deep circular reactor, Rondelez *et al*. were able to demonstrate DNA-based travelling prey-predator reaction-diffusion fronts. Recently, this work has been elaborated[237] to a more gen-

[b]Polymerase is a general term used to denote an enzyme that synthesises a DNA strand using nucleotides as building blocks.

[c]A nicking enzyme produces a break in one strand of a double-stranded DNA, leading to its unwinding.

[d]An exonuclease cleaves (*i.e*. removes through hydrolysis) one nucleotide at a time from a polynucleotide chain.

eral method for achieving control over the reaction and diffusion parameters of DNA components employed in programmable reaction-diffusion networks.

5.3 Replicating systems within reaction-diffusion environment

The notion of competitive autocatalysis in reaction-diffusion systems was investigated[239] by Showalter *et al*. using simulations. The authors examined the possibility of achieving complete selectivity in product formation in a simplified system composed of two autocatalytic processes competing for a shared building block **A** (**Equation 5.1** and **Equation 5.2**).

$$\mathrm{A} + \mathrm{B} \rightarrow 2\mathrm{B} \quad \text{rate} = k_{\mathrm{B}}[\mathrm{A}][\mathrm{B}] \tag{5.1}$$

$$\mathrm{A} + \mathrm{C} \rightarrow 2\mathrm{C} \quad \text{rate} = k_{\mathrm{C}}[\mathrm{A}][\mathrm{C}] \tag{5.2}$$

The autocatalytic products, **B** and **C**, each have a specific diffusion coefficient ((D_{B} and D_{C}) and their formation from their constituent components (reaction of **A** and **B**, or **A** and **C**) is governed by a specific rate constant (k_{B} and k_{C}). Using computer simulations, Showlater *et al*. showed[239] that the selectivity observed in a propagating reaction-diffusion front depends on the relative ratios of the rate constants ($k_{\mathrm{C}}/k_{\mathrm{B}}$) and the diffusion coefficients ($D_{\mathrm{B}}/D_{\mathrm{C}}$) specific for the two competing replicators—**B** and **C**. **Figure 5.8** shows an example simulation, where the two autocatalytic species have an identical diffusion coefficient ($D_{\mathrm{B}} = D_{\mathrm{C}}$) and the rate constant k_{B} is twice as high as k_{C}. The outcome of these simulation (**Figure 5.8**) shows that the species with the higher rate constant, **B**, as the kinetically favoured product, forms a wave with constant velocity. In contrast, product **C**, travels from the initial seeding point almost exclusively through diffusion, propagating over a significantly smaller distance.

In a situation where the diffusion coefficients are identical, the rate constant for the formation of the competing species plays a crucial role in determining the selectivity. This straightforward example of competition between **B** and **C** was elaborated into a simulation where the two species are more closely matched in replicating abilities. More specifically, Showlater *et al*. showed that when $D_{\mathrm{B}}/D_{\mathrm{C}} > k_{\mathrm{C}}/k_{\mathrm{B}}$, the autocatalytic template **B** forms a propagating wave selectively (favoured by diffusion) while if $D_{\mathrm{B}}/D_{\mathrm{C}} < k_{\mathrm{C}}/k_{\mathrm{B}}$, the shared component **A** gets converted predominantly to product **C**, favoured kinetically, instead.

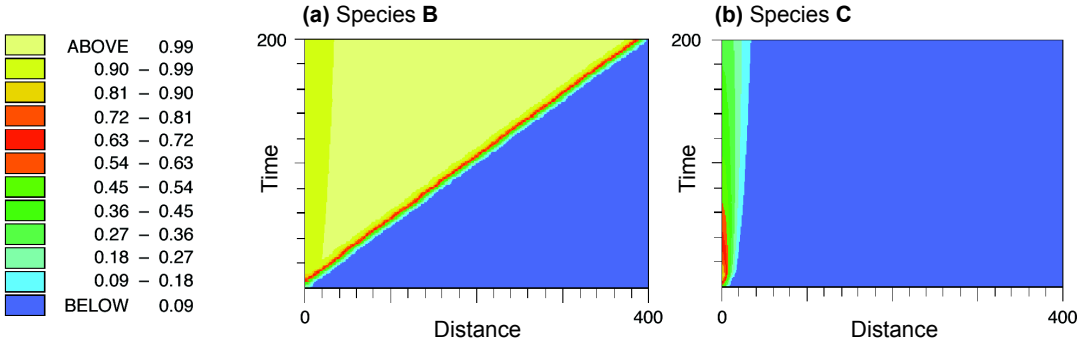


Figure 5.8 Simulated time *vs* distance travelled plots for the modelled evolution of autocatalytic species **(a) B** and **(b) C** when $D_B/D_C = 1$ and $k_C/k_B = 0.5$. Concentration ranges of each product formed are represented by the colour key on the left, where blue colour represents the lowest and yellow the highest concentration. Figure adapted from Ref. 239 with permission from The Royal Society of Chemistry.

This theoretical study showed that reaction-diffusion format can offer better selectivity in a network of competing autocatalytic reactions than a well-stirred batch reactor system, where the final composition always comprises a mixture of products. The degree of selectivity in a system of competing autocatalytic reactions, coupled to diffusion, can be fine-tuned by altering the reaction and diffusion parameters of the reacting species within a particular system. Despite the clear potential utility of the discussed study as a general model, the simulations employed in this work examine the formation of autocatalytic products, **B** and **C**, as a simple second order reaction with respect to the reaction components, **A** and **B** or **C** (**Equation 5.1** and **Equation 5.2**). This approach differs markedly from that of the minimal model of self-replication, introduced in **Chapter 1**, where a template molecule is required to establish an autocatalytic cycle (**Equation 5.3**), with an overall reaction order of up to 3. Additionally, the study[239] by Showalter *et al*. does not take into consideration any crosscatalytic interactions and thus the simulated results are not wholly applicable to a network of replicators where such pathways operate—including the **T1**–**T2** network examined in this thesis.

$$\mathrm{A} + \mathrm{B} + \mathrm{T} \rightarrow 2\mathrm{T} \quad \text{rate} = k_1[\mathrm{A}][\mathrm{B}][\mathrm{T}] \tag{5.3}$$

5.3.1 Synthetic replicating systems in reaction-diffusion environment: objectives and previous work

Far-from-equilibrium reaction conditions present a reaction format that permits replicating systems to overcome the fundamental limit on selectivity and emergence of complex behaviour that can develop in networks under closed conditions, driven by kinetic selection—*i.e*. WSBR format. In particular, replicating systems in nature never

operate under equilibrium conditions—instead, they direct their own formation within a dynamic, yet kinetically-stable,[24,39,240] mixture of components that constantly interact with the outside world. The rich variety of behaviour exhibited by systems in propagating reaction-fronts described thus far suggests that selective formation of one species is a phenomenon that does not need to be exclusively confined to biological organisms. Therefore, in order to progress towards such complexity in behaviour in small-molecule based synthetic networks, it is crucial to move away from the study of molecules under equilibrium conditions—and the environment of a propagating reaction-diffusion front could provide a suitable environment for such studies. This chapter will explore the possibility of applying and adapting the known, well-studied network of replicators **T1**–**T2** examined in **Chapter 3** and **Chapter 4** into a system that can be studied under far-from-equilibrium conditions.

For several year now, the Philp laboratory has been working[241] towards developing a replicating system that could function within the reaction-diffusion environment—a system capable of initiating and sustaining a chemical propagating reaction-diffusion front. Exploiting the design of a well-established efficient replicating system, first reported[127] by Kassianidis and Philp in 2006, Bottero and co-workers developed[241] a replicating system with an optical signature that permits monitoring of the progress of the replication progress within the reaction-diffusion environment in real time (**Figure 5.9**).

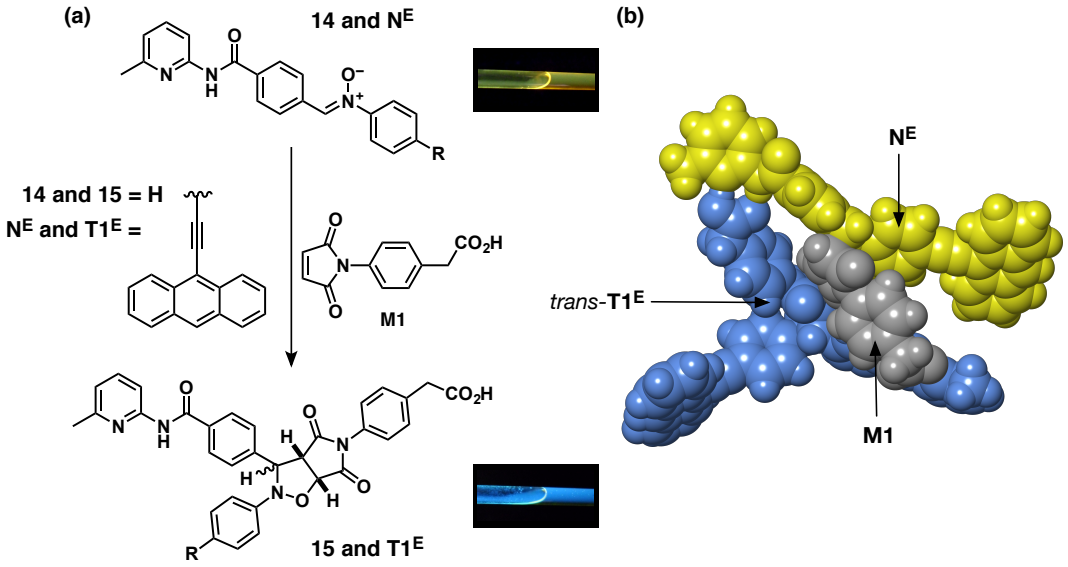


Figure 5.9 **(a)** Design of a self-replicating system (R = H), *trans*-**15**, produced by the reaction of a **14** and maleimide **M1**. Replacement of the R group within the molecular framework of *trans*-**15** replicator with a 9-ethynylanthracene optical tag, produces replicator *trans*-$\mathbf{T1^E}$. Formation of replicator *trans*-$\mathbf{T1^E}$ from nitrone $\mathbf{N^E}$ and **M1** is associated with a change in fluorescence from bright yellow (nitrone) to blue (cycloadduct). **(b)** Calculated (RM1) space-filling structure of the transition state [$\mathbf{N^E}$·**M1**·$\mathbf{T1^E}$] allowing the template-directed formation of *trans*-$\mathbf{T1^E}$ replicator. Figure adapted with permission from Ref. 241. Copyright 2016 American Chemical Society.

The original replicating system[e] (**Figure 5.9a**), based on the reaction between an aryl-phenyl nitrone **14** and maleimide **M1** was altered to contain a 9-ethynylanthracene fluorescent tag (**Figure 5.9a**) in place of the fluorine, to give nitrone $\mathbf{N^E}$. Computational analysis of the redesigned system showed that the template formed through the reaction of $\mathbf{N^E}$ and **M1**, $\mathbf{T1^E}$, should be capable of self-replication through the ternary catalytically-active complex [$\mathbf{N^E}$·**M1**·$\mathbf{T1^E}$] (**Figure 5.9b**).

Nitrone $\mathbf{N^E}$, equipped with an anthracene tag, exhibits bright yellow fluorescence when irradiated with a long-wavelength UV light, which undergoes a dramatic change to blue fluorescence upon its reaction with maleimide **M1**. The ability of the redesigned system to self-replicate was confirmed through a series of kinetic experiments. Reaction of an equimolar solution (10 mM) of the fluorescent nitrone $\mathbf{N^E}$ with **M1** at 0 °C in $CDCl_3$, resulted in efficient production of $\mathbf{T1^E}$, with extremely high diastereoselectivity for the *trans* diastereoisomer ([*trans*]/[*cis*] ratio of cycloadducts *ca.* 100). Addition of 10 mol% of preformed *trans*-$\mathbf{T1^E}$ template to the reaction components necessary for its formation shortened the lag period dramatically (**Figure 5.10a**), confirming that the system maintains its ability to self-replicate despite the presence of the significantly more bulky anthracene tag.

The reliance of the efficient self-replicating system on molecular recognition was confirmed further by analysis of the reaction of the fluorescent nitrone with a recognition-disabled maleimide **M3** (**Figure 5.10b**). Reaction of these two components at 0 °C in $CDCl_3$ resulted in significantly lower conversion to *trans*-$\mathbf{T1^E_{dis}}$ and an increase in the formation of the *cis*-$\mathbf{T1^E_{dis}}$ diastereoisomer, producing these two products in ratio of 3 after five hours (**Figure 5.10b**), a value characteristic for systems capable of exploiting the bimolecular pathway only. Kinetic fitting of the experimental data for replicator *trans*-$\mathbf{T1^E}$ revealed a kinetic effective molarity of 16.2 M—a value higher than $EM_{kinetic}$ determined for **T2** replicator (9.47 M) but lower than the effective molarity of **T1** replicator ($EM_{kinetic} = 64.0$ M, **Chapter 3**).

[e]The design of a self-replicating system used as the basis for construction of a replicator with suitable optical signature of replication is very similar to the **T1** replicator examined in this thesis, however, the original design exploits a non-fluorinated aryl-phenyl group on the nitrone (**14**).

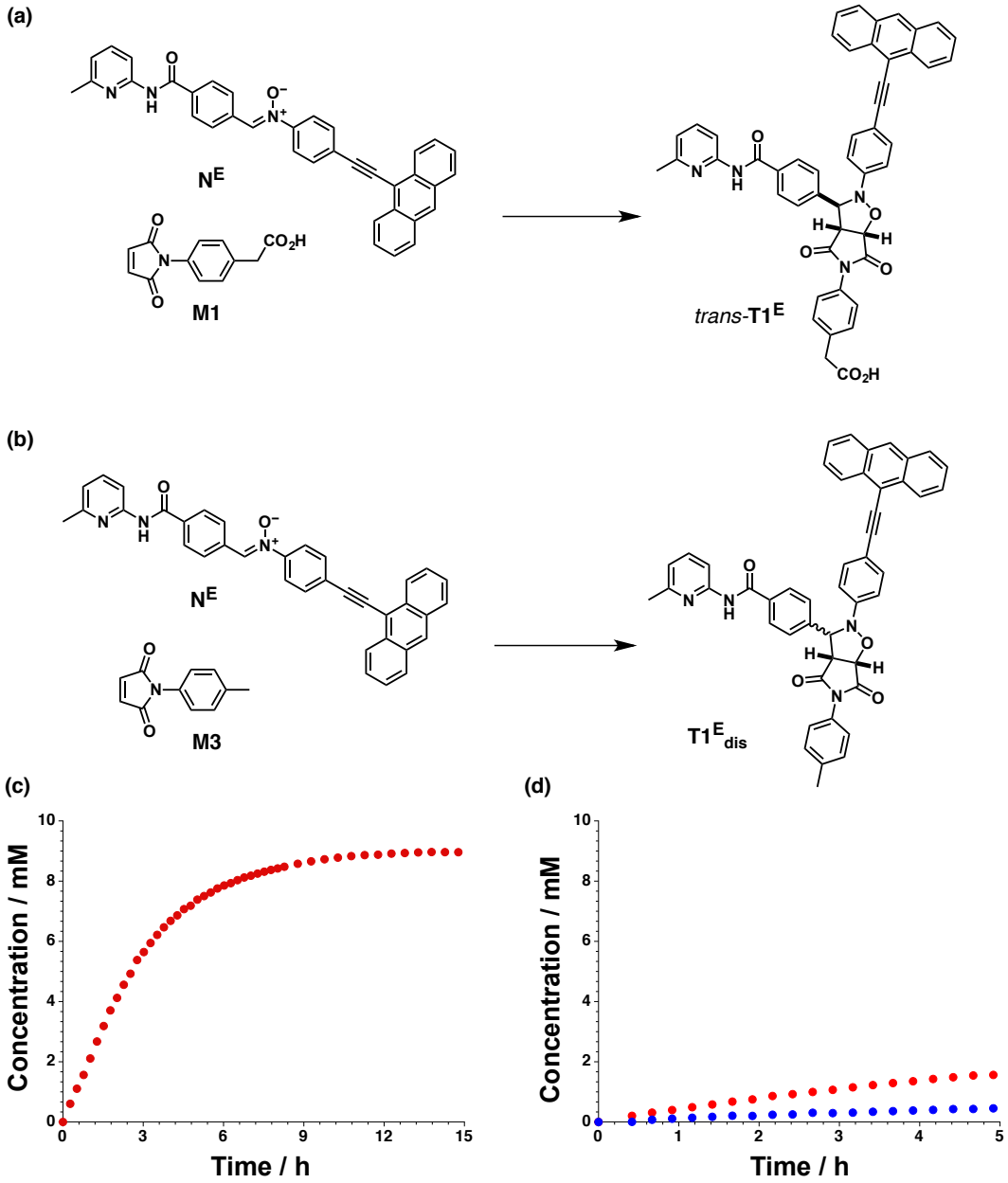


Figure 5.10 **(a)** A self-replicating system, *trans*-$\mathbf{T1}^{\mathbf{E}}$ is formed by the reaction of nitrone $\mathbf{N}^{\mathbf{E}}$ equipped with a fluorescent anthracene tag with a phenylacetic acid maleimide **M1**. Reaction of these components in the presence of 10 mol% of preformed *trans*-$\mathbf{T1}^{\mathbf{E}}$ template results in the disappearance of the lag period in the **(c)** concentration *vs* time profile, confirming the ability of the system to self-replicate. **(b)** Reaction of nitrone $\mathbf{N}^{\mathbf{E}}$ with maleimide **M3**, lacking the carboxylic acid recognition site, shows that the 1,3-dipolar cycloaddition proceeds **(d)** slowly and with low diastereoselectivity, affording both *trans* and *cis* diastereoisomers. Figure adapted with permission from Ref. 241. Copyright 2016 American Chemical Society.

The change in fluorescence, associated with the formation of replicator *trans*-**T1**$^{\mathbf{E}}$ was confirmed as the signature of the underlying autocatalytic reaction by examining the replicating system using UV-vis spectroscopy, at the same reaction temperature and concentration as employed in the ^{1}H NMR kinetic experiments. Specifically, aliquots of a reaction mixture containing **N**$^{\mathbf{E}}$ and **M1** were periodically taken, diluted to 100 µM and the UV-Vis absorption of each sample recorded (**Figure 5.11c**).

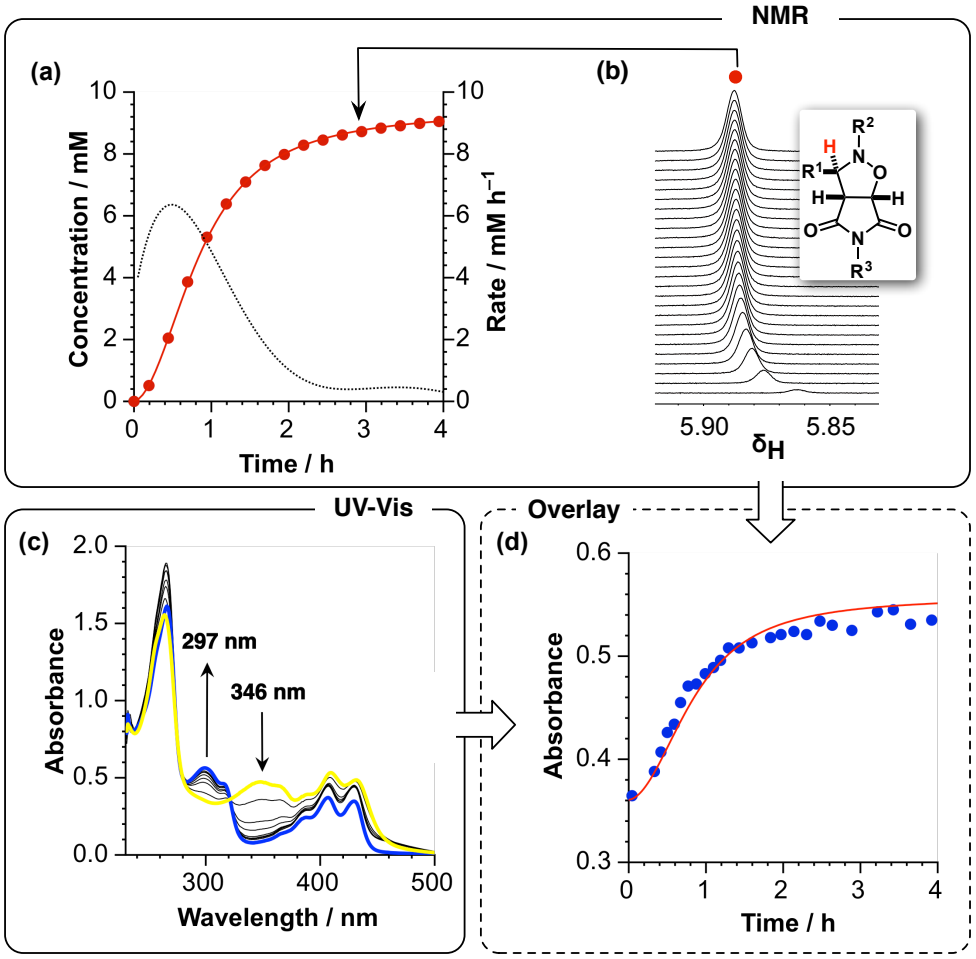


Figure 5.11 (a) Concentration (red circles) and rate (black dotted line) *vs* time profile determined for the reaction of **N**$^{\mathbf{E}}$ and **M1** (10 mM) by ^{1}H NMR spectroscopy (499.9 MHz) at 20 °C in $CDCl_3$. The best fit of the data to a kinetic model for the reaction is depicted in full red line. **(b)** Stack of of partial arrayed ^{1}H NMR spectra showing the resonance, and the progressive change in its chemical shift, corresponding to the *trans* proton on the isoxazolidine unit in *trans*-**T1**$^{\mathbf{E}}$ formed during the reaction of **N**$^{\mathbf{E}}$ and **M1**. **(c)** UV-Vis spectra recorded during the reaction of the same nitrone and maleimide components, showing the changes to the absorbtion over the course of the reaction. **(d)** Comparison of the absorbance at 297 nm *vs* time (blue circles represent experimental data from panel **c**, whereas red line shows the computed absorbance determined using the best fit of the kinetic model to the NMR data shown in **a**). Figure adapted with permission from Ref. 241. Copyright 2016 American Chemical Society.

The collected spectra showed clear changes in the absorption over the course of the reaction—a decrease in the intensity of the band at 346 nm was observed to occur

simultaneously with an increase in the intensity of the bands at 264 and 297 nm. In order to compare the reaction profile determined by UV-vis spectroscopy to the data obtained in the ^{1}H NMR kinetic experiment, the authors used the reaction and association parameters obtained through the processes of kinetic fitting of the NMR data for the formation of replicator **T1^{E}**, to compute the expected absorbance at 297 nm—thereby reconstructing the reaction profile based on data predicted by UV-vis spectroscopy. **Figure 5.11d** illustrates the excellent agreement between the kinetic data obtained by UV-vis spectroscopy and the computed expected absorbance profile, based on the best fit of NMR kinetic data to a kinetic model (**Figure 5.11a**) showed an excellent agreement.

Using UV-vis and ^{1}H NMR spectroscopy, Bottero *et al*. have established unambiguously the replication efficiency of *trans*-**T1^{E}** replicator. Generally, formation of a propagating reaction-diffusion front can be initiated by addition of a small amount of a solution of preformed autocatalyst to an unstirred solution of the unreacted material at a specific location. Once initiated, the autocatalytic species can propagate away from the site of addition—*i.e*. from the region of high concentration. Deuterated chloroform, the reaction solvent employed in the formation of replicator *trans*-**T1^{E}**, is highly volatile, and, thus, the ability of the designed system to initiate and sustain a propagating reaction-diffusion front was assessed using the environment of a 50 µL microsyringe, as opposed to a flat plate. The experimental design, employing two gas-tight syringes, is illustrated in **Figure 5.12a** and **Figure 5.12b**.

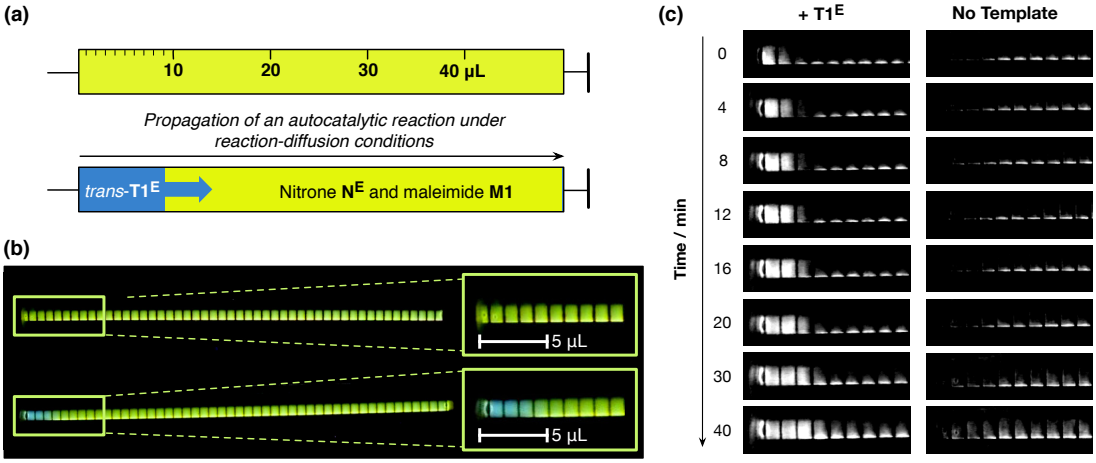
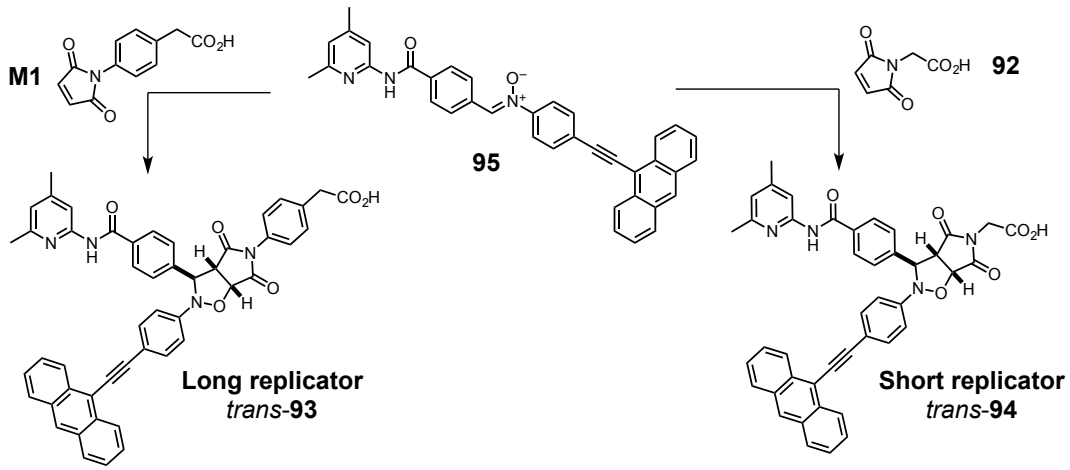


Figure 5.12 **(a)** Graphical and **(b)** photographic representation of the experimental set-up for the examination of a synthetic replicator under reaction-diffusion conditions using two 50 µL syringe. Top syringe represents the control condition in each case, lacking any preformed replicator template. The bottom syringe is pre-filled with unreacted nitrone and maleimide, to which a small amount of solution of preformed replicator is added at one end. **(c)** Processed grey-scale images collected over time, using a 365 nm UV lam to illuminate a template instructed reaction-diffusion experiment (left-hand column, + *trans*-**T1^{E}**) and control condition, lacking template (right-hand column, no template). Figure adapted with permission from Ref. 241. Copyright 2016 American Chemical Society.

One 50 µL syringe was completely filled with a 5 mM solution of $\mathbf{N^E}$ and **M1** (top syringe, **Figure 5.12a** and **b**) whilst the second syringe, pre-filled with the same building blocks contained also 2 µL of 10 mM solution of *trans*-$\mathbf{T1^E}$, added at one end of the syringe. Examination of the processed images acquired by illumination of both syringes at a sequence of time points with a 365 nm UV light (**Figure 5.12c**) showed that addition of preformed replicator template solution allowed *trans*-$\mathbf{T1^E}$ to establish a propagating reaction-diffusion front. In contrast, no front was visible in the experiment lacking the preformed template, and, instead, the syringe underwent a uniform change in fluorescence, from yellow to blue. This experimental demonstration of a reaction-diffusion front driven by a small-molecule organic synthetic replicator, described by the Philp laboratory opens up the possibilities of exploring networks of interconnected replicators within a reaction-diffusion environment, *i.e.* under far-from-equilibrium conditions that could allow such systems to express the phenomenon of selective replication.

The proof-of-principle simple fluorescent replicator was extended[203] in currently unpublished work, to include a second maleimide[f] **92**, lacking the phenyl spacer, thereby producing a system of two replicators, referred to as the long replicator **93** and short replicator **94** (**Scheme 5.1**), competing for a shared nitrone building block **95**, equipped with a 4,6-dimethylamidopyridine recognition unit.



Scheme 5.1 Replicating network comprised of two autocatalytic replicators termed as long replicator **93** and short replicator **94**, formed by the reaction of nitrone **95** with **M1** and a short maleimide **92**.

The extended system analysed for the first time a propagating reaction-diffusion front comprising simultaneously not one, but two replicators. Initially, the kinetic profiles of each replicator were analysed in isolation, within the environment of a NMR

[f]The short maleimide **92** employed for the synthesis of replicator **94** is identical to maleimide **7a** mentioned previously in **Chapter 1**.

tube, followed by analysis within an competition environment where both maleimides could react with the nitrone component. The kinetic analysis of the long and short replicator by ^{1}H NMR spectroscopy (all components at 5 mM, at 4 °C, in $CDCl_3$) in isolation showed that the long template *trans*-**93** is markedly more efficient at templating its own formation than the short template *trans*-**94**—82% and 39% conversion after 16 hours, respectively (**Figure 5.13a** and **b**). Further kinetic experiments established that while both replicators are capable of self-replication, neither of the crosscatalytic cycles is active in the network (**Figure 5.13c**). As suggested by the outcome of the simulation study reported by Showalter *et al.*, the strictly autocatalytic nature of the network means that the more efficient template should be able to outcompete the weaker one, assuming similar diffusion coefficients (the replicators differ by a single phenyl ring only).

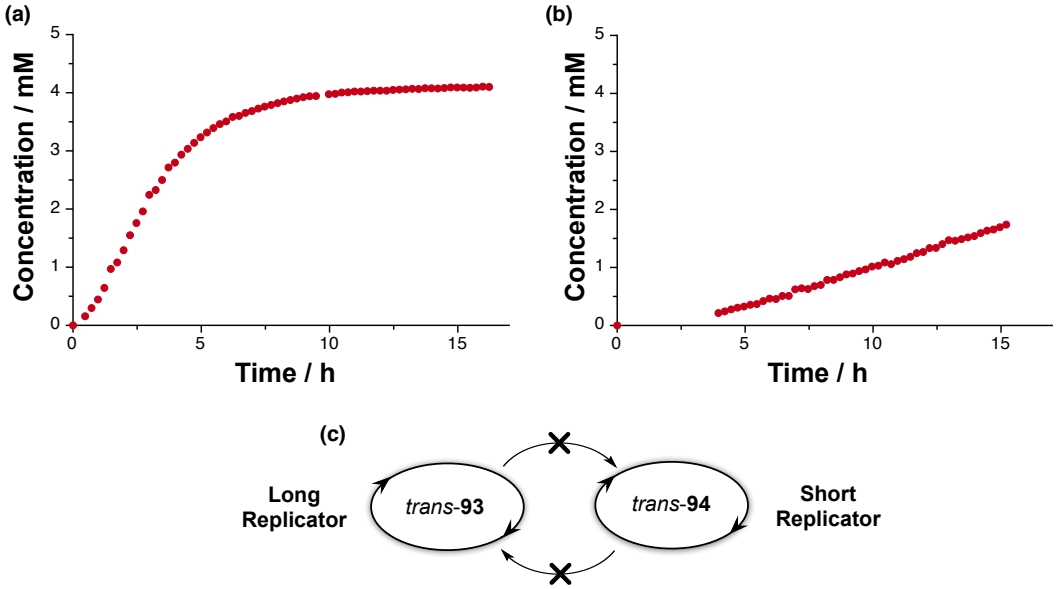


Figure 5.13 Concentration *vs* time profile for the reaction of nitrone **95** with **(a) M1** maleimide to make *trans*-**93** and **(b)** short maleimide **92** to produce *trans*-**94** ([**95**] = [**M1**] or [**92**] = 5 mM, 4 °C, $CDCl_3$) as determined by ^{1}H NMR (499.9 MHz) spectroscopy. **(c)** The catalytic relationships in the network of *trans*-**93** and *trans*-**94**. Figure adapted from Ref. 203.

In fact, when the two replicators were examined under the same reaction conditions within a competition scenario (**Figure 5.14a**), the reaction outcome revealed a strong bias towards the formation of the long replicator *trans*-**93**. Nevertheless, this significantly more efficient replicator was unable to outcompete the weaker *trans*-**94** completely, resulting in a ratio of [*trans*-**93**]/[*trans*-**94**] of 3.7 after two days. Next, the formation of the two replicators was examined in the presence of 10 mol% of preformed template of the long (**Figure 5.14b**) and the short replicator (**Figure 5.14c**). In these instructed competition kinetic experiments, the ratio of [*trans*-**93**]/[*trans*-**94**] after two days increased from 3.7 to 5 in the presence of the long template *trans*-**93** and decreased to 2.48 when instructed with the short template *trans*-**94**.

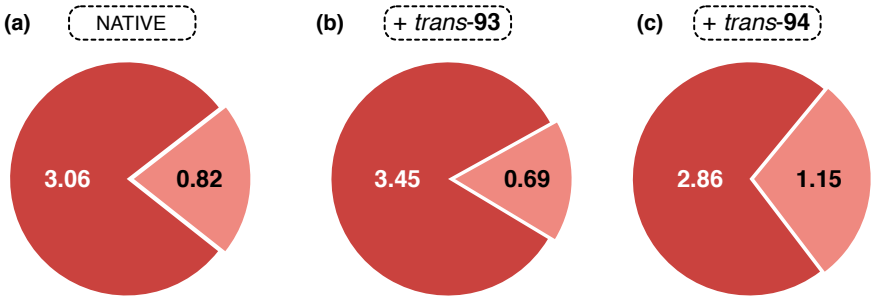


Figure 5.14 Pie-chart representation of the outcome of competition kinetic experiments between a long maleimide **M1** and short maleimide **92 (a)** in the absence of template, **(b)** instructed with long template *trans*-**93** and **(c)** with short replicator *trans*-**94**. Concentration of each product formed has been corrected for the template added and is shown in mM. Figure taken from Ref. 203.

Both the short and the long replicators were designed to contain the previously described 9-ethynylanthracene optical tag, which was previously[241] found to impart the replicating system with an optical signature that permits the reaction progress to be monitored in real time, by illumination under UV. An example evolution of a *trans*-**93** replicator driven reaction-diffusion front over time, formed by addition of a small amount (5 µL) of preformed template solution to a syringe already preloaded with nitrone **95** and **M1** is shown in **Figure 5.15**.

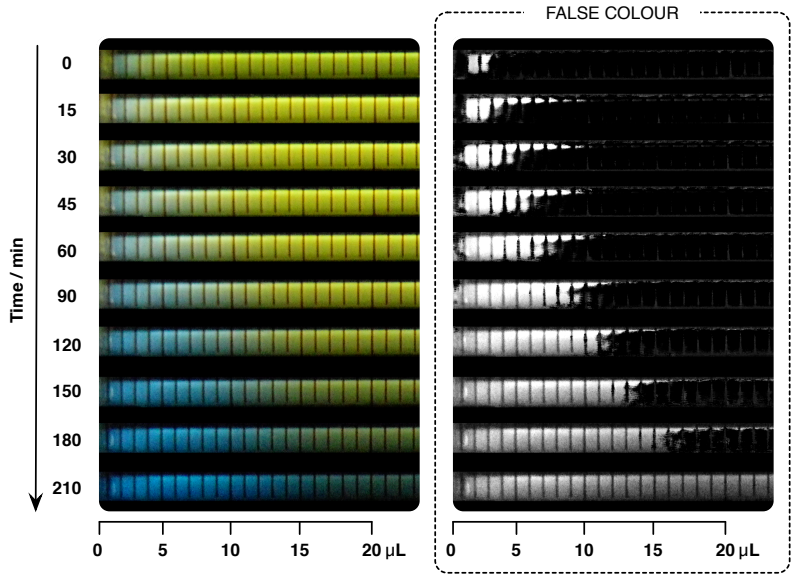


Figure 5.15 Time-lapse images demonstrating the propagation of a reaction-diffusion front, formed by addition of *ca.* 5 µL of solution of long maleimide *trans*-**93** to a syringe prefilled with 45 µL of **M1** and **95**, over time (2.5 mM, $CDCl_3$, 4 °C). Time-lapse photographs on the right have been processed for clarity. Figure adapted from Ref. 203.

Comparison of the the long and short replicator in reaction-diffusion environment revealed that the long template forms a reaction-diffusion front that propagates noticeably faster than the front arising from the short template. The two replicators were

further examined in a competition scenario, where a 50 µL gas-tight syringe was filled with 45 µL of equimolar solution containing all three unreacted reaction components ([**M1**] = [**95**] = [**92**] = 2.5 mM, $CDCl_3$, 4 °C), followed by injection of additional 5 µL of preformed long template *trans*-**93** ([*trans*-**93**] = 2.5 mM). Analysis of the syringe contents, corresponding to the length of the formed wave by ^{1}H NMR spectroscopy, revealed the presence of the resonance arising from the long *trans*-**93** cycloadduct replicator only (**Figure 5.16**).

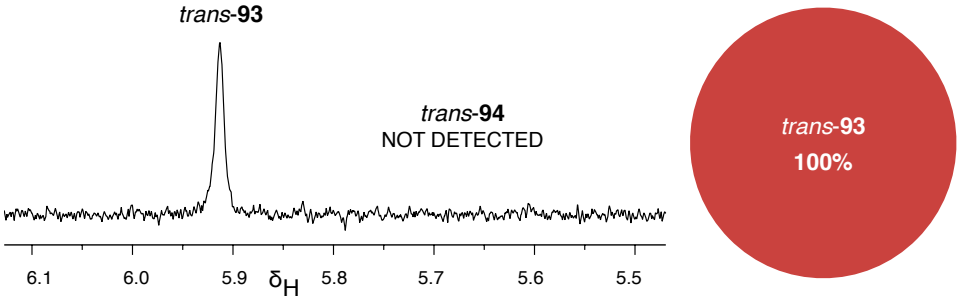


Figure 5.16 Partial ^{1}H NMR spectrum portraying the *trans* resonance arising from the exclusive product detected in the competition between a recognition-enabled maleimide **M1** and **92** for the nitrone **95** within a reaction-diffusion system, instructed with *trans*-**93** (2.5 mM, $CDCl_3$, 4 °C). Figure taken from Ref. 203.

Comparison of the selectivity obtained in the competition network within a closed system to selectivity within a reaction-diffusion front suggests that complete selectivity in a network comprised of two relatively simple organic replicators might be possible. The examined system, however, investigated the competition between two replicators that are fairly unevenly-matched in their kinetic abilities. Additionally, as pointed out in regards to the theoretical study by Showalter *et al.*, this particular network of a short and a long replicator contains no active crosscatalytic channels, and, thus, the outcome of the competition is fairly straightforward. The effect of performing the competition experiments within the reaction-diffusion format using more interconnected networks, *i.e.* those involving crosscatalysis (such as the **T1**–**T2** network), on the selectivity for a particular replicator, remains to be investigated. Lastly, while the NMR analysis of the reaction-diffusion front instructed by the long template showed the presence of a single product only—a result highly suggestive of complete selectivity for the template *trans*-**93**—it was not possible to establish unambiguously whether the resonance corresponding to *trans*-**93** in the ^{1}H NMR spectrum (**Figure 5.16**) represents the template added to initiate the formation of the front only, or also the template formed within the reaction-diffusion front. At this stage, therefore, it is not completely clear whether the single resonance associated with *trans*-**93**, detected in the competition

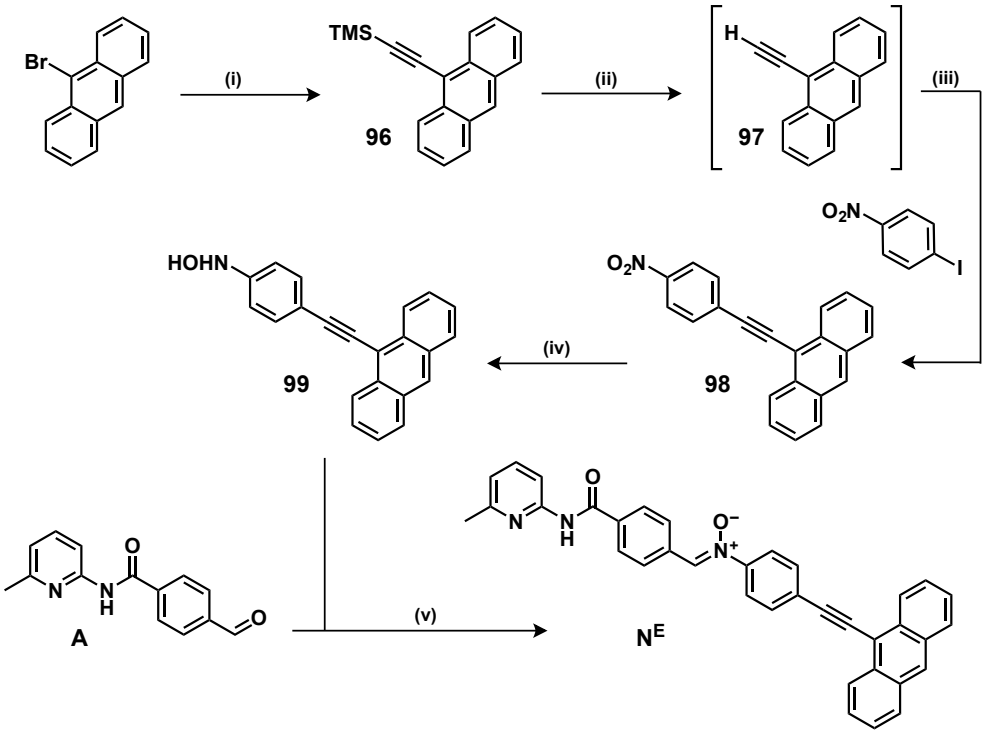
reaction-diffusion experiment (**Figure 5.16**), represents a demonstration of a selective replication.

5.4 A two-replicator network within a reaction-diffusion environment

Examination of the **T1**–**T2** network under reaction-diffusion conditions—*i.e.* an environment where the replicators **T1** and **T2** could overcome the limit imposed on selectivity for one replicator over another by kinetic selection—required that the replicators in this network be equipped with a fluorescent tag. To this end, the anthracene tag employed successfully by the Philp laboratory in the previous work was also exploited for the monitoring of the reaction-diffusion front in the **T1**—**T2** network—the aryl-fluorine tag on the nitrone was substituted with the 9-ethynyl anthracne, producing the key fluorescent nitrone, referred to as $\mathbf{N^E}$ (**Scheme 5.2**).

5.4.1 Synthesis of nitrone equipped with an optical tag

Starting from a commercially available 9-bromoanthracene, the first step in the synthetic route towards the fluorescent nitrone $\mathbf{N^E}$ (**Scheme 5.2**) was a palladium catalysed Sonogashira coupling of a TMS-acetylene group to give the anthracene product **96**, bearing a protected alkyne. The TMS intermediate **96** was subsequently deprotected using potassium carbonate in THF/MeOH, to afford the terminal alkyne **97** in a quantitative yield—alkyne **97** was used in the next step directly without purification. A nitrophenyl functional group was introduced *via* another Sonogashira coupling of the **97** with 4-bromonitrobenzene. The key nitro intermediate **98** was reduced to the corresponding hydroxylamine **99** using hydrazine monohydrate and a rhodium on carbon catalyst (5%wt), leaving the triple bond intact under these conditions. In the final step, the crude hydroxylamine **99** was condensed with aldehyde **A** in a mixture of EtOH/$CHCl_3$, producing the fluorescent nitrone $\mathbf{N^E}$. Syntheses of aldehyde **A** and both maleimides, **M1** and **M2**, were described in **Chapter 3**.



Scheme 5.2 Reagents and conditions employed in the synthesis of fluorescent nitrone $\mathbf{N^E}$. (i) TMS-acetylene, CuI, PPh_3, $PdCl_2(PPh_3)_2$, Et_3N, 80 °C, 18 h, 85%. (ii) K_2CO_3, MeOH/THF, RT, 3 h, quantitative conversion. (iii) PPh_3, $PdCl_2(PPh_3)_2$, Et_3N, 55 °C, 66 h, 19%. (iv) Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, THF, RT, 25 min, quantitative conversion. (v) EtOH/$CHCl_3$, dark, 5 °C, 3 days, 46%.

5.4.2 Experimental set-up for the observation of reaction diffusion fronts

Kinetic analyses of the fluorine tagged nitrone $\mathbf{N^F}$ with maleimides **M1** and **M2** was reported in detail in **Chapter 3**. Fluorescent nitrone $\mathbf{N^E}$ is equipped with the same reactive nitrone site, with the same relative orientation to the recognition unit to that in the fluorine tagged nitrone $\mathbf{N^F}$. The presence of an anthracene tag within a replicating system has been shown[203,241] to exert little effect on the efficiency of replication in the previous work within the Philp laboratory, discussed in the preceeding sections. As a result of this similarity of $\mathbf{N^F}$ to the target fluorescent nitrone $\mathbf{N^E}$, exhaustive kinetic analyses of the fluorescent version of the replicators, namely $\mathbf{T1^E}$ and $\mathbf{T2^E}$, was not undertaken at this point.

As a first step in the examination of whether the replicators in the interconnected $\mathbf{T1^E}$–$\mathbf{T2^E}$ network can form a reaction-diffusion front with enhanced selectivity, the capacity of each replicator to establish a propagating chemical wave, in response to seeding with solution of preformed template was examined in isolation. In a typical wave experiment, a 50 µL gas-tight syringe was filled first with 50 µL of an equimolar

solution containing the maleimide and $\mathbf{N}^{\mathbf{E}}$ ([**M1**] or [**M2**] = [$\mathbf{N}^{\mathbf{E}}$] = 5 mM). In instructed experiments, 5 µL of the syringe contents were ejected and instead 5 µL of a preformed autocatalytic template, specific for each experimental condition, were carefully taken up into the syringe, with extreme care to avoid introduction of air bubbles that could interfere with the front propagation. The general experimental set-up, employed for the analysis of propagating reaction-diffusion fronts is illustrated in **Figure 5.17**.

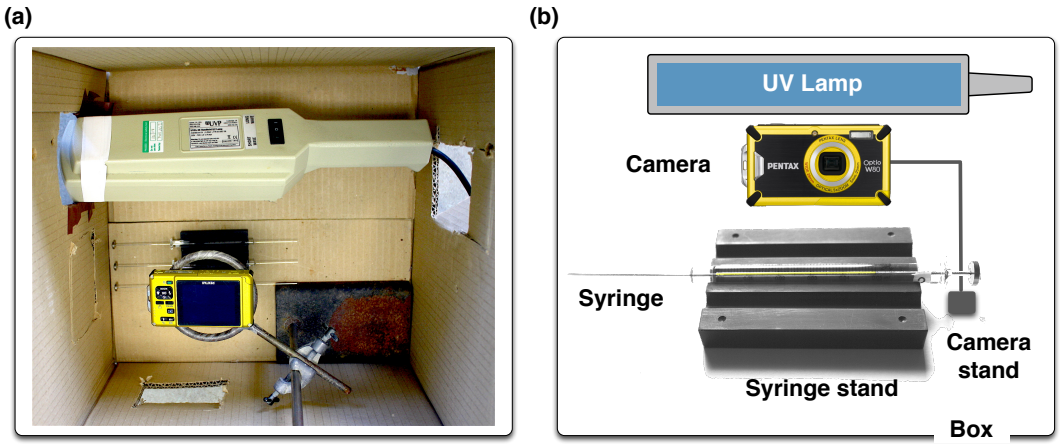


Figure 5.17 **(a)** Image and **(b)** cartoon representation of the experimental set-up employed in the reaction-diffusion experiments, showing a 50 µL gas-tight microsyringe in a custom-made syringe stand and the approximate position of the camera and the UV lamp.

As a result of the limited solubility of the fluorescent templates, *trans*-$\mathbf{T1}^{\mathbf{E}}$ and *trans*-$\mathbf{T2}^{\mathbf{E}}$, at the required concentrations (10 to 20 mM), the preformed templates employed in the seeding process were synthesised using the non-fluorescent nitrone $\mathbf{N}^{\mathbf{F}}$. Each experiment was carried out[g] in $CDCl_3$, at 5 °C, and the evolution of the system under reaction-diffusion conditions was monitored by illumination with a 365 nm UV light and time-lapse photography.

5.4.3 Analysis of replicators under reaction-diffusion conditions

5.4.3.1 Uninstructed reaction-diffusion experiments

In order to demonstrate that a reaction-diffusion front does not emerge spontaneously in this $\mathbf{T1}^{\mathbf{E}}$–$\mathbf{T2}^{\mathbf{E}}$ under the chosen experimental conditions, a control experiment was performed where no solution of preformed template was added to the microsyringe. Instead, the entire volume of the 50 µL syringe was filled with a stock solution containing the fluorescent nitrone $\mathbf{N}^{\mathbf{E}}$ and the two maleimides (**M1** and **M2**) ([**M1**] = [**M2**] = [$\mathbf{N}^{\mathbf{E}}$] = 5 mM). The changes observed in the uninstructed syringe experiment over time

[g]The same temperature at which the **T1**–**T2** network was examined in **Chapter 3** and **Chapter 4**. The temperature was maintained at a stable temperature by undertaking the experiments in a temperature regulated cold room, whose temperature was regulated at 5 °C.

were monitored by time-lapse photography and illumination under 365 nm UV light (**Figure 5.18**).

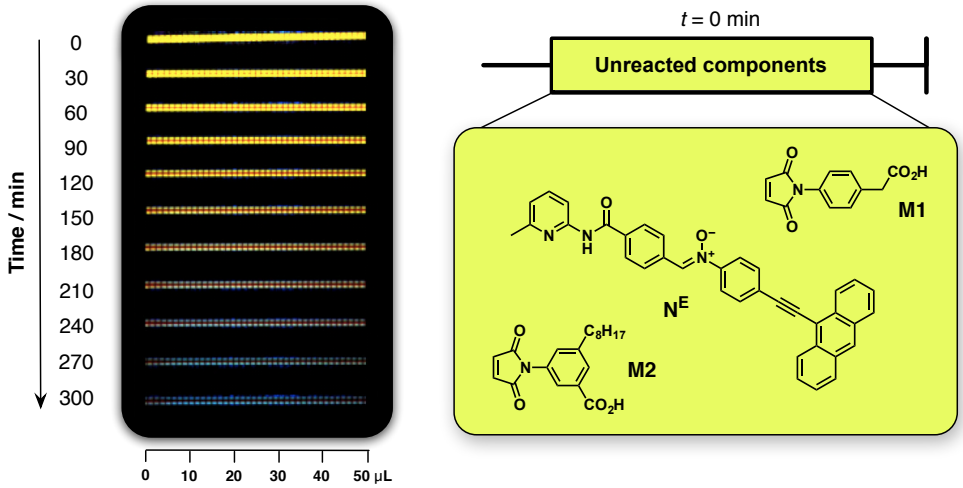


Figure 5.18 Control uninstructed reaction-diffusion experiment: a 50 µL syringe was filled with an equimolar solution of $\mathbf{N}^{\mathbf{E}}$ and maleimides **M1** and **M2** (all components at 5 mM, in $CDCl_3$) and the syringe was monitored over time at 5 °C, by time-lapse photography. Data taken from Ref. 242.

The control reaction-diffusion experiment contain no preformed template added, only the nitrone and maleimide components. Therefore, the blue colour associated with the formation of the cycloadducts is expected to occur evenly throughout the syringe over time, as the reaction progresses. Indeed, examination of the control reaction-diffusion experiment (**Figure 5.18**) shows that blue fluorescence develops throughout the syringe despite the absence of any added preformed template. The blue colour is present homogeneously throughout the entire length of the syringe, confirming that the blue colour is not the result of a chemical propagating wave. Examination of the syringe contents by NMR spectroscopy revealed that the blue fluorescence in the control experiment is a result of two processes: (i) the formation of cycloadducts through the slow bimolecular pathway and later the template-mediated autocatalytic pathway and (ii) the slow gradual decomposition of the fluorescent nitrone $\mathbf{N}^{\mathbf{E}}$ as a result of extended exposure to UV light. Decomposition of the nitrone to its constituent elements, *i.e.* the hydroxylamine and aldehyde, is also accompanied by slow formation of the corresponding azoxy adduct (formation of azoxy species was first discussed in the context of dynamic covalent libraries in **Chapter 4**). In order to minimise the extent and the rate of decomposition processes in the replicating systems examined within the reaction-diffusion experiments, illumination under UV should be reduced as much as possible and, most certainly, not used continuously. In particular, experiments intended for qualitative analysis should only be exposed to UV light as a means of establishing the extent of front propagation at a particular time-point.

5.4.3.2 Template-instructed reaction-diffusion experiments

The formation of a propagating reaction-diffusion front by both the **T1** and **T2** replicators has been first examined in isolation, *i.e*. conditions where only a single maleimide is present within the unreacted reaction mixture in the syringe. In each experiment, the syringe was filled with 45 µL of unreacted nitrone $\mathbf{N^E}$ and one maleimide (**M1** or **M2**, 5 mM). The front formation was initiated by the addition of 5 µL of solution of pre-formed fluorine-tagged template of either $\mathbf{T1^F}$ or $\mathbf{T2^F}$ (10 mM). Both **T1** (**Figure 5.19**) and **T2** (**Figure 5.20**) replicators were shown to be capable of initiating propagating reaction-diffusion fronts fronts in isolation within a 50 µL syringe[h], as evidenced by examination of the distance *vs* time travelled by each front over the course of the experiment.

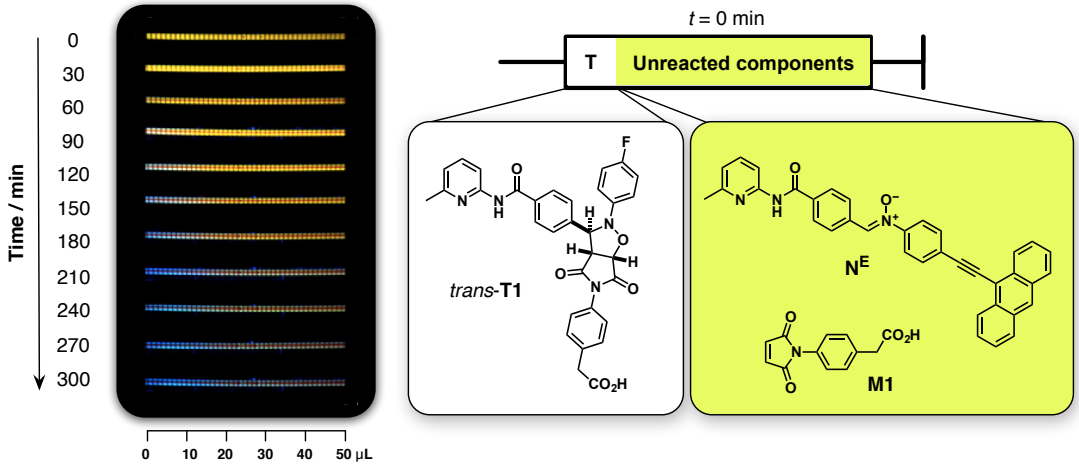


Figure 5.19 Propagating reaction-diffusion $\mathbf{T1^E}$ front initiated by addition of *trans*-**T1** (5 µL, 20 mM) to a 50 µL syringe pre-filled with an equimolar solution of $\mathbf{N^E}$ and **M1** (45 µL, all components at 5 mM, in $CDCl_3$) and the reaction progress was monitored at 5 °C, by time-lapse photography. Data taken from Ref. 242.

[h]In order to ensure that the 50 µL syringe is the most suitable size-format for examination of reaction-diffusion fronts, additional control experiments were carried out in both a 25 and a 100 µL gas-tight syringes. However, attempts to initiate a propagating reaction-diffusion front in either the 25 or 100 µL syringes proved unsuccessful. These experiments confirmed that upon transferring the replicating system to larger syringe, the degree of convective mixing becomes too extensive whereas the wave propagation slows down too significantly within the smaller syringe.

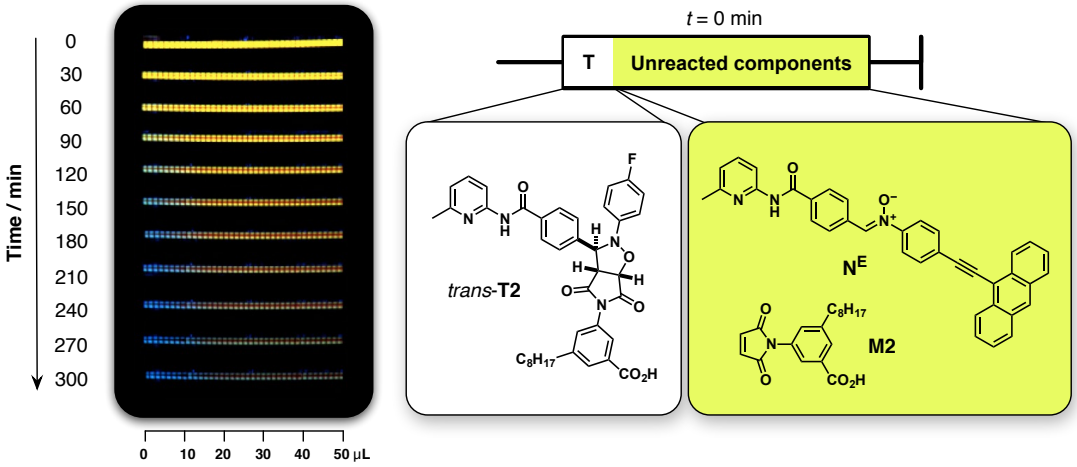


Figure 5.20 Propagating reaction-diffusion $\mathbf{T2}^{\mathbf{E}}$ front initiated by addition of *trans*-**T2** (5 µL, 20 mM) to a 50 µL syringe pre-filled with an equimolar solution of $\mathbf{N}^{\mathbf{E}}$ and **M2** (45 µL, all components at 5 mM, in $CDCl_3$) and the reaction progress was monitored at 5 °C, by time-lapse photography. Data taken from Ref. 242.

The reaction-diffusion fronts in these isolated experiments were initiated using the fluorine-tagged preformed templates and it was noted during these experiments that the clear solution of the preformed templates in $CDCl_3$ mixed with the highly fluorescent nitrone solution after injection. In order to test the extent of initial mixing (after injection), 5 µL of a solution containing fluorescent template was added to 45 µL of clear $CDCl_3$ (already present in the syringe). The experiment showed that the region where convective mixing takes place extends to *ca.* to 7 µL mark in the syringe. Despite the initial mixing, examination of **Figure 5.19** and **Figure 5.20** revealed that addition of the preformed templates initiated successfully a propagating wave of $\mathbf{T1}^{\mathbf{E}}$ and $\mathbf{T2}^{\mathbf{E}}$, respectively. After five hours, these templates have reached *ca.* 20 µL for $\mathbf{T1}^{\mathbf{E}}$ and 15 µL for $\mathbf{T2}^{\mathbf{E}}$. The slightly lower distance propagated by $\mathbf{T2}^{\mathbf{E}}$ replicator relative to $\mathbf{T1}^{\mathbf{E}}$ is in agreement with the kinetic profiles of these two replicators in isolation, reported in **Chapter 3**.

5.4.4 Competition within the reaction-diffusion environment

Analysis of the **T1**–**T2** network under competition conditions in both a closed (**Chapter 3**) and dynamic environment (**Chapter 4**) showed that in the absence of instruction, the **T2** replicator reaches a slightly higher concentration than the **T1** replicator. If the network of replicators is instructed with preformed **T1** template, **T1** is formed preferentially and the resulting [**T2**]/[**T1**] ratio is lower than the ratio in the uninstructed experiment, with no template added—a behaviour that can be described by the IF Boolean logic (**Table 5.1**).

Table 5.1 Molecular IF logic gate describing the outcome of instructed competition experiment in the **T1**–**T2** replicating network. 1 signifies the presence of preformed template added to the competition experiment and also an increase in the replicator ratio: [**T2**]/[**T1**]. Only presence of **T1** template results in a decrease in the [**T2**]/[**T1**] ratio.

IF		
T1	**T2**	**[T2] / [T1]↓**
0	**0**	**0**
1	**0**	**1**
0	**1**	**0**
1	**1**	**1**

In the next step, the two replicators were examined under competition conditions within the reaction-diffusion environemnt—in the presence of no template, with **T1**, with **T2**, and with both templates added simultaneously to seed the front formation. It was envisaged that the addition of either the preformed template of **T1** or **T2** to the microsyringe containing a fluorescent nitrone $\mathbf{N^E}$ and both maleimides, **M1** and **M2**, would allow the replicator capable of establishing a more efficient propagating front to extinguish the slower reaction-diffusion front over time. A graphical representation showing the proposed gradual increase in selectivity is illustrated on the example of a *trans*-**T2** instructed reaction-diffusion system (**Figure 5.21**).

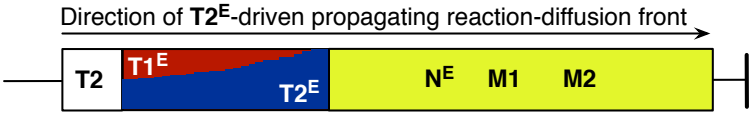


Figure 5.21 Graphical representation of the gradual extinguishing of a $\mathbf{T1^E}$ by a $\mathbf{T2^E}$ reaction-diffusion front in a competition experiment instructed by fluorine-tagged *trans*-**T2** replicator.

The uninstructed control experiment described in **Figure 5.18** served as a useful indicator of how long a reaction mixture containing only two maleimides and a nitrone takes to exhibit blue fluorescence throughout the syringe. This control experiment can also be used as a baseline for comparison of the ratio of cycloadducts formed within template-instructed reaction-diffusion experiments. In order to examine the ratio of products formed within the four devised reaction-diffusion experiments, the syringe contents corresponding to the length of the wave[i] were transferred into a micro NMR tube (**Figure 5.22**) and were diluted with fresh $CDCl_3$ as necessary, before being examined by 499.9 MHz ^{1}H NMR spectroscopy.

[i] If the length of the front could not be established unambiguously, a sample of 15 µL from the respective syringe was analysed by ^{1}H NMR spectroscopy.

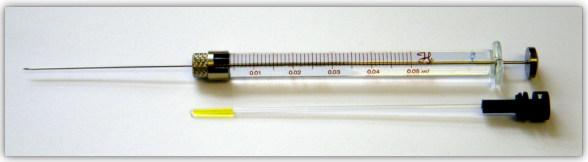


Figure 5.22 Example of a 50 µL syringe and a 1.7 nm NMR tube employed during ^{1}H NMR spectroscopic analysis.

Periodic monitoring of the wave progress under UV revealed that a wave seeded by **T1** (20 mM) formed a propagating wave. The waves initiated by **T2** (20 mM) and both **T1** and **T2** (10 mM each) extended over a smaller distance in the syringe than the **T1**-instructed front, and it was not possible to establish unambiguously how far the front has propagated.

Analysis of the control uninstructed experiment (**Figure 5.23a**) by 499.9 MHz ^{1}H NMR spectroscopy was straightforward, showing only two resonances, corresponding to $\mathbf{T1^{E}}$ and $\mathbf{T2^{E}}$ observed in the region of the ^{1}H NMR spectrum associated with the *trans* cycloadducts (6.2–5.6 ppm). Results of the three instructed, competition-wave samples (**Figure 5.23b** to **d**), however, proved to be difficult to interpret. The estimations of the level of selectivity in these reaction-diffusion competition experiments was challenging particularly because of the possibility that some of the template formed in the background bimolecular reaction could have been introduced into the examined sample by removing inadvertently a volume that was larger than the propagation length of the formed front.

In these three instructed experiments, the non-fluorescent templates *trans*-**T1** and *trans*-**T2** were used to seed the reaction-diffusion fronts—giving rise to up to four different resonances specific to **T1**, $\mathbf{T1^{E}}$, **T2** and $\mathbf{T2^{E}}$ in the ^{1}H NMR spectra. The ^{1}H NMR spectra determined by analyses of the **T1**-instructed (**Figure 5.23b**) and **T2**-instructed (**Figure 5.23c**) reaction-diffusion experiments showed the presence of a single resonance only, in the region associated with the preformed fluorine-tagged templates **T1** and **T2** (5.9–5.7 ppm). Curiously, only one resonance belonging to the preformed template was observed in the ^{1}H NMR spectrum of the wave sample instructed with both templates. A potential rationale for this finding can be found in the changes in the chemical shift of the *trans* resonance arising from the **T1** template, as a function of **T2**—behaviour that was discussed in detail in **Chapter 3**, suggesting that the presence of only a single resonance can be explained by the overlap of **T1** and **T2** in the experiment shown in **Figure 5.23d**. In two out of the three instructed experiments (**Figure 5.23b** and **Figure 5.23d**), more than a single signal was found in the *trans* region specific for the formed templates. The analysis of these two competition experiments was further complicated by the changes in the chemical shifts of the

products formed. More specifically, both of the peaks thought to belong to **T1**$^{\mathbf{E}}$ and **T2**$^{\mathbf{E}}$ had a chemical shift that was different from those found in the control experiment. Again, these changes are very similar to the variation in chemical shift exhibited by the **T1**–**T2** network during the competition kinetic experiments in a closed system in **Chapter 3**. A single resonance, associated with the fluorescent *trans* cycloadduct formed the reaction-diffusion front was observed in ^{1}H NMR spectrum of the **T2**-instructed experiment only (**Figure 5.23c**). This finding suggests that while selectivity might have improved relative to the uninstructed experiment undertaken in a closed reactor, complete selectivity for one replicator over another was attained only in the **T2**-instructed reaction-diffusion front.

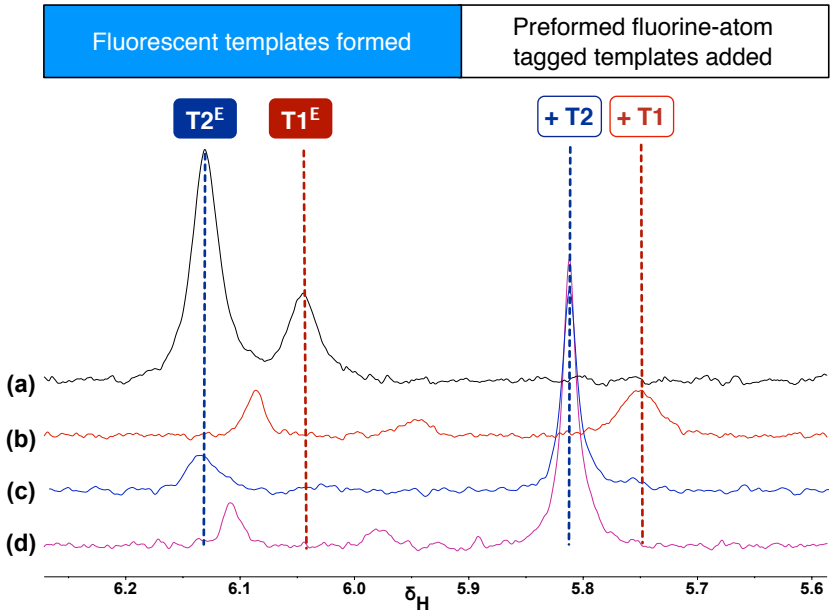


Figure 5.23 Overlay of partial ^{1}H NMR spectra (499.9 MHz) of competition waves initiated by (a) no template, (b) **T1** added (20 mM), **(c) T2** added (20 mM), **(d) T1** and **T2** (10 mM each) added. Each syringe was filled with 45 µL of 5 mM solution of **N**$^{\mathbf{E}}$, **M1** and **M2** in $CDCl_3$. Waves were run at 5 °C. Data taken from Ref. 242.

The presented results from these reaction-diffusion experiments examining the **T1**$^{\mathbf{E}}$–**T2**$^{\mathbf{E}}$ network under competition conditions demonstrated that, whilst the employed reaction format may increase the selectivity for one replicator over another in this network connected by a shared building block, the current experimental set-up, and in particular the low reaction volume available for analysis using micro NMR tubes, introduces uncertainty in to the interpretation of the results determined within the reaction-diffusion environment.

5.5 Towards improved analysis of replicating networks within reaction-diffusion environment

The previous work in the Philp laboratory showed that a single self-replicating system exploiting 1,3-dipolar cycloaddition reaction as the template-forming step in a closed system reaction format can be analysed in real-time using UV-Vis spectroscopy in addition to the ^{1}H or ^{19}F NMR spectroscopy analysis employed more traditionally, with comparable results. The environment of a 50 µL gas-tight syringe was found to be a suitable vehicle for examining formation and propagation of reaction-diffusion fronts in synthetic self-replicating systems under model far-from-equilibrium conditions. Successful initiation of propagating reaction-diffusion fronts was achieved by addition of a small amount of solution containing preformed replicator template at a specific location within the syringe. Examination of the competition reaction-diffusion experiments, however, showed that the transfer of a replicating system into a reaction-diffusion environment, where autocatalysis is coupled to size-dependent diffusion, presents an analytical challenge.

Following the challenges arising in the analysis of the presented reaction-diffusion experiments, several conditions were identified that are crucial for successful analysis of a propagating reaction-diffusion fronts:

1. It should be possible to monitor the reaction progress visually.
2. It must be possible to distinguish between the preformed template added to initiate the front formation and the template formed within the reaction-diffusion front.
3. A spectroscopic method, suitable for the analysis of the small quantities of material that comprise the contents of a syringe in a reaction-diffusion experiment, needs to be identified or a method for increasing the sample volume developed.
4. If possible, the distance travelled by a propagating reaction-diffusion front should be compared to a control diffusion experiment (one containing only the nitrone reactive component within the syringe).

Fulfilment of condition 2 is essential for removing any ambiguity associated with using a template that is formed within the reaction also as the preformed template solution. Each of these conditions will be discussed briefly, followed by the experimental changes and steps that can be undertaken to ensure that these conditions are satisfied.

The reaction sample within the reaction-diffusion front is inhomogeneous, and, therefore, it is not possible to take out even a small sample volume of the reaction material for analysis throughout the course of the wave experiment—doing so would disturb

the concentration gradient and the front propagation. Nevertheless, the anthracene tag employed in the previous experiments allows the reaction progress to be monitored and assessed visually (by illumination using a long wavelength UV light), as the fluorescence changes dramatically upon transformation of the starting materials in to products—fulfilling the first condition.

Spectroscopic techniques such as UV-Vis in particular, have been shown to afford results comparable to those obtained by NMR spectroscopy for monitoring the progress of the 1,3-dipolar cycloaddition reaction in a closed system. However, these techniques have limited capacity to distinguish between different cycloadducts with similar fluorescence properties. Currently, only NMR spectroscopy provides a means of unambiguously distinguishing different cycloadduct species, and ^{19}F NMR spectroscopy, with its large range of chemical shifts presents a particularly suitable method. In order to be able to assess the ratio of products formed within a competition experiment in a reaction-diffusion front by NMR spectroscopy, it is necessary to be able to distinguish between the added template and that formed within the reaction-diffusion front. The capacity to distinguish between the templates added and formed necessitates that these templates can be distinguished spectroscopically. The only way for the resonances arising from these templates to be different, is to employ a component possessing a slightly altered fluorescent tag. This tag needs to maintain its suitable fluorescent properties, yet the chemical of the resonance it corresponds to must be sufficiently different as to be able to distinguish it from the other resonances already present in the system. To this extent, a number of anthracene tags have been developed, several equipped with a fluorine tag. The presence of a fluorine atom tag is particularly useful as it facilitated analysis of the small quantities of material available by ^{19}F NMR spectroscopy and allowed observation of resonances arising from the unreacted nitrone as well as the formed templates. The availability of the same reaction components equipped with slightly different fluorescent tags means that a different tag can be employed for the preformed template than for the one present on the unreacted material—satisfying condition 2.

The reaction-diffusion experiments are carried out in a 50 µL gas tight syringe, providing only very small quantities of material for analysis at the end of each experiment. Therefore, analysis of the contents of each reaction-diffusion front has been challenging, requiring either specialised equipment, such as a 1.7 mm micro NMR tube (only suitable for analysis by ^{1}H NMR spectroscopy) and prolonged experimental time (the acquisition of even a relatively simple ^{1}H NMR spectrum can take over an hour) as a result of high sample dilution, needed to reach sufficient sample volume. In order to increase the quantity of material available from the wave experiments for

analysis, several replicas of a desired experiment can be performed simultaneously and the contents of these identical fronts combined for analysis—the increased volume can allow samples to be analysed using non-specialised 5 mm NMR tubes, employing ^{19}F NMR spectroscopy, satisfying condition 3.

Using the so-called batch-method for performing reaction-diffusion experiments, the problem of small sample quantity can be addressed. In the previous **T1^{E}** and **T2^{E}** competition wave experiment (**Figure 5.23**), each reaction-diffusion front was analysed as a single sample. In fact, in this experiment, the remains of the syringe (*i.e.* where the front propagation was not observed) were not analysed at all. As suggested in the graphical representation showing a hypothetical gradual increase in selectivity in a *trans*-**T2** instructed reaction-diffusion system (**Figure 5.21**), the composition of the reaction-diffusion front might change over time (as a function of distance travelled by the front). Therefore, more information about a reaction-diffusion front can be gained by analysis of the entire 50 µL gas-tight syringe, divided into a number of sections. For example, the simplest way of processing the syringe contents would be to divide it into five equal sections, A to E, each comprising 10 µL (**Figure 5.24a**). Nevertheless, this division method might not provide the most effective means of analysis a reaction-diffusion front—the highest selectivity for one replicator over another can be expected within the reaction-diffusion front formed in the sections closer to the location of the added preformed template, as opposed to the sections far away. This expectation stems from the fact that the components in the syringe will start reacting as soon as the unreacted nitrone and maleimide building blocks are mixed, producing the replicator templates evenly throughout the syringe—which means that a propagating reaction-diffusion front is competing for the reagents with these so-called 'background' reactions. A better division system, illustrated in **Figure 5.24b**, can be envisaged where section A (5 µL) contains mostly the preformed template and also a small fraction of the templates formed within the reaction-diffusion front. Sections B and C examine 10 µL each, revealing the changes in the front contents that have reached the 15 µL and 25 µL mark. Section D examines the syringe contents from 25 µL to 50 µL—section that will contain mostly templates formed through the background reactions. Running each reaction-diffusion experiment simultaneously within multiple syringes, and the subsequent examination of the syringe contents in several sections should permit the evolution of a propagating reaction-diffusion front to be determined by NMR spectroscopy.

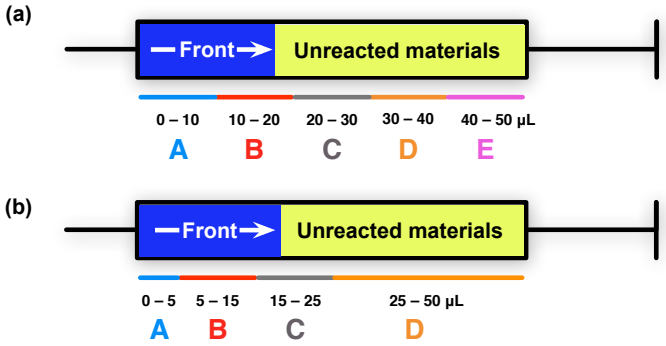


Figure 5.24 Graphical representations showing how a 50 µL gas-tight syringe can be divided in to **(a)** five and **(b)** four different sections for separate analysis by NMR spectroscopy.

Thus far, a method for increasing sample volume and an approach to analysis of front progression have been discussed. Another challenging aspect associated with the analysis of experiments undertaken within a reaction-diffusion environment is the need to ensure that the contents of each sample/section are analysed as quickly as possible (ideally within a similar time window) in order to avoid the continuing reaction processes within the system from affecting the outcome of analysis. The most efficient reaction processes in replicating systems are mediated by molecular recognition—while it is not possible to stop the bimolecular, recognition-independent reaction pathways completely, it is possible to exploit the addition of a small quantity of DMSO-d_6 to each sample in order to disrupt hydrogen bonding. The DMSO added interferes with the recognition-mediated processes in the system, thereby slowing down the reaction progress to the rate of the bimolecular reactions. This procedure allows the recognition processes in each sample/section to be stopped at the same time, after which the prepared samples can be analysed qualitatively by ^{1}H NMR and/or ^{19}F NMR spectroscopy.

An understanding of the behaviour of replicating systems within the reaction-diffusion environment, and thus also the ability to design a system with suitable reaction and diffusion properties, requires knowledge of how much do the individual reaction and diffusion parameters contribute to the front propagation. In order to be able to deconvolute the contribution of diffusion to the front propagation, control diffusion experiments can be undertaken. Namely, the preformed template can be added to a syringe containing only the fluorescent nitrone component—a situation where no new products can be formed within the mixture, and any changes in fluorescence within the syringe can be attributed to diffusion. These control experiments can employ various volumes of preformed fluorescent template—allowing the distance travelled by a particular template over time to be assessed visually and compared to the distance obtained for a propagating reaction-diffusion front, addressing condition 4, identified previously.

5.5.1 Second generation of optical tags for use in reaction-diffusion experiments

An ideal optical tag, suited for analysis of replicating systems within the environment of a reaction-diffusion front needs to be sufficiently soluble, ideally possess a fluorine tag and exhibit suitable optical properties. For example, whilst the 4-fluorophenyl tag employed in **Chapter 3** and **Chapter 4** possessed sufficient solubility properties, it had a clear shortcoming in that it lacks the desired optical features that would enable the nitrone and template products to be distinguished visually. Nevertheless, despite the absence of a fluorescent tag, 4-fluorophenyl group can be employed as the preformed template used to instruct the reaction and initiate front formation. In contrast to the 4-fluorophenyl tag, the original fluorescent nitrone $\mathbf{N^E}$ is equipped with an anthracene tag that exhibits the optimal optical properties, but lacks a fluorine tag and exhibits limited solubility at the template stage, most likely as a result of the triple bond present within the tag structure. Inspired by the original structure of the anthracene tag employed in $\mathbf{N^E}$, several optical tags were designed (**Figure 5.25**) for use in reaction-diffusion experiments. In an attempt to improve the solubility of $\mathbf{N^E}$, the triple bond was deleted from the design, affording a nitrone $\mathbf{N^A}$. In order to facilitate analysis by ^{19}F NMR spectroscopy, the core structures of tags $\mathbf{N^E}$ and $\mathbf{N^A}$ were adapted to incorporate a fluorine atom, producing four fluorescent tags, referred to as: $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$, $\mathbf{N^{A10F}}$ and $\mathbf{N^{AFPh}}$, at the nitrone stage (**Figure 5.25**).

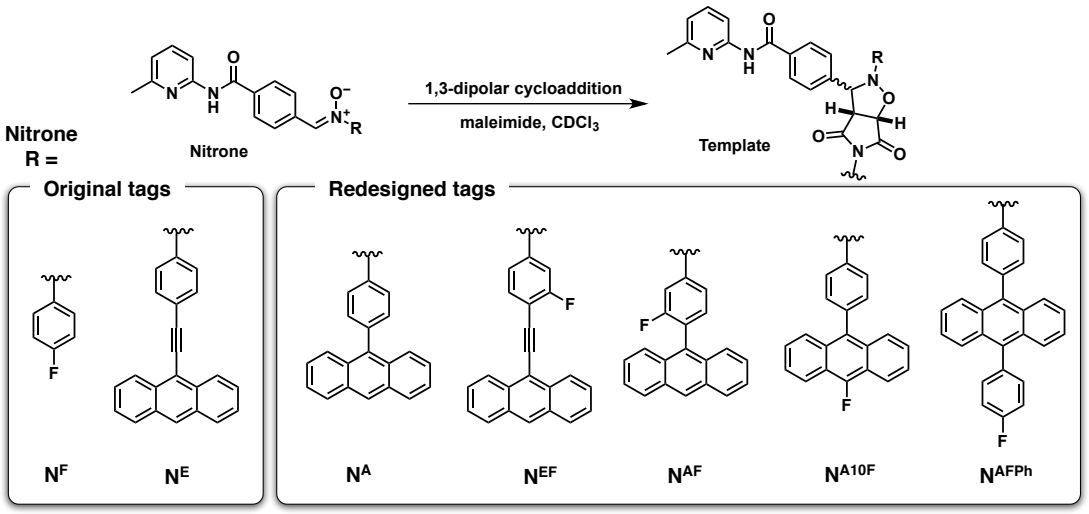
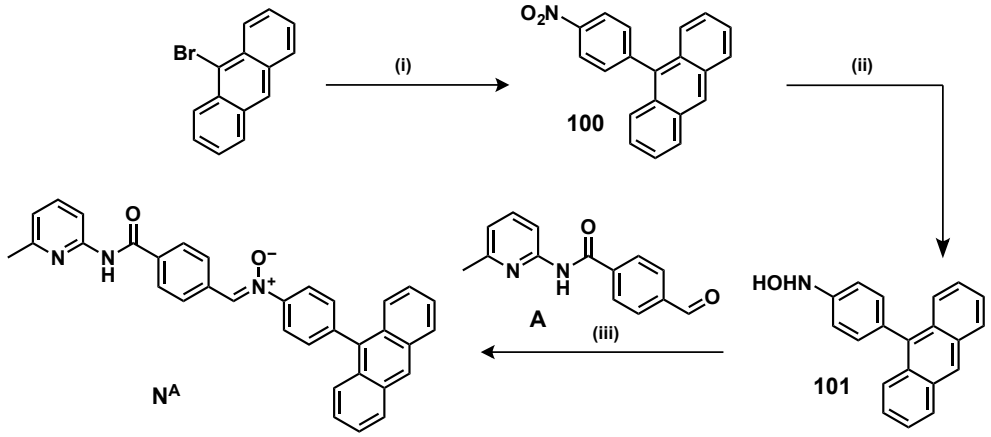


Figure 5.25 A nitrone reacts with maleimide through 1,3-dipolar cycloaddition to afford a self-replicating template. R corresponds to the structure of the original aryl-fluorine and anthracene tags employed thus far and those designed to improve solubility and permit analysis by ^{19}F NMR spectroscopy. The label corresponding to the nitrone incorporating each tag is shown below each structure. Labels of the nitrones incorporating the redesigned tags denote the position of the fluorine atom within the structure or its absence: $\mathbf{N^A}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$, $\mathbf{N^{A10F}}$ and $\mathbf{N^{AFPh}}$.

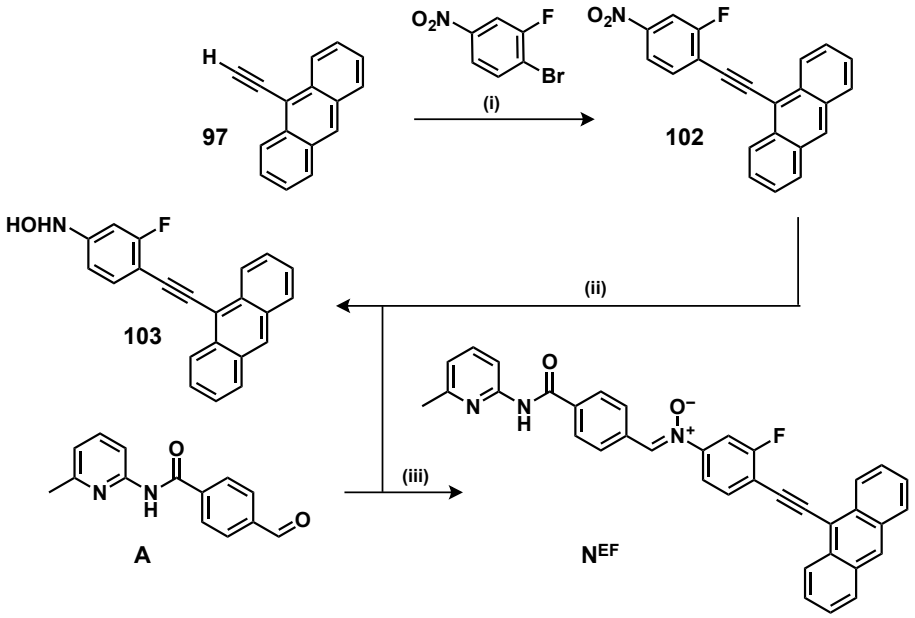
5.5.1.1 Syntheses of new optical tags

The initial step in the synthesis of nitrone $\mathbf{N^A}$ (**Scheme 5.3**) exploited a Suzuki coupling of 9-bromoanthracene and 4-nitrphenylboronic acid, to give the key nitro intermediate **100**, in a good 79% yield. The nitro compound **100** was reduced to the corresponding hydroxylamine **101** using the rhodium on carbon catalyst (5% wt) and hydrazine monohydrate. Condensation of hydroxylamine **101** with aldehyde **A** synthesised previously afforded the desired nitrone $\mathbf{N^A}$ (44%), exhibiting yellow fluorescence under long wavelength UV light.



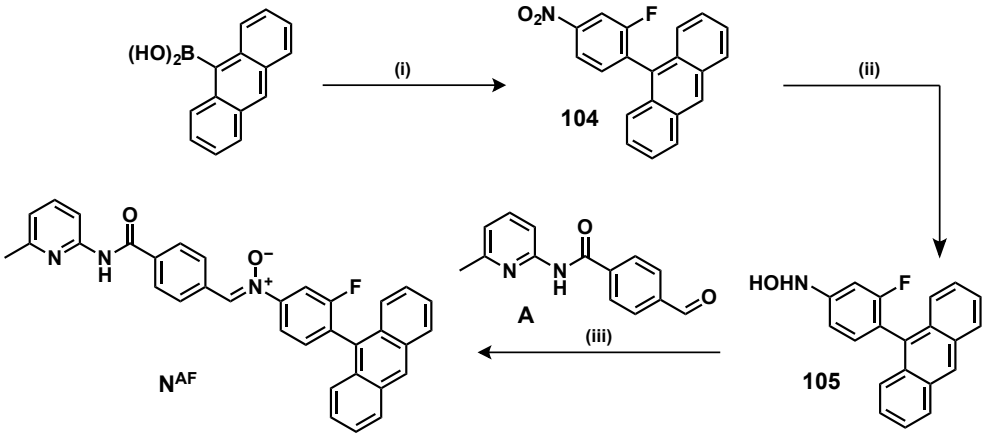
Scheme 5.3 Reagents and conditions employed in the synthesis of Nitrone $\mathbf{N^A}$. (i) 4-nitrobenzeneboronic acid, $Pd(PPh_3)_4$, K_3PO_4, PhMe/EtOH/H_2O, 110 °C, 1 day, 79%. (ii) THF, Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, RT, 20 minutes, quantitative conversion. (iii) EtOH/$CHCl_3$, dark, 3 days, 4 °C, 44%.

Synthesis of the fluorinated version of the original 9-ethynylanthracene nitrone $\mathbf{N^E}$, $\mathbf{N^{EF}}$ (**Scheme 5.4**), commenced from terminal aklyne **97**, also employed in the synthesis of $\mathbf{N^E}$ (**Scheme 5.2**). Sonogashira coupling of the anthracene alkyne **97** with 4-bromo-3-fluoro-nitrobenzene produced the key nitro intermediate **102** (26%). Reduction of **102** using hydrazine monohydrate and rhodium on carbon (5% wt), afforded the hydroxylamine **103** in essentially quantitative yield, without affecting the ethynyl group. Condensation of the hydroxylamine **103** with aldehyde **A** in a mixture of EtOH/$CHCl_3$ afforded the desired nitrone $\mathbf{N^{EF}}$ (37%).



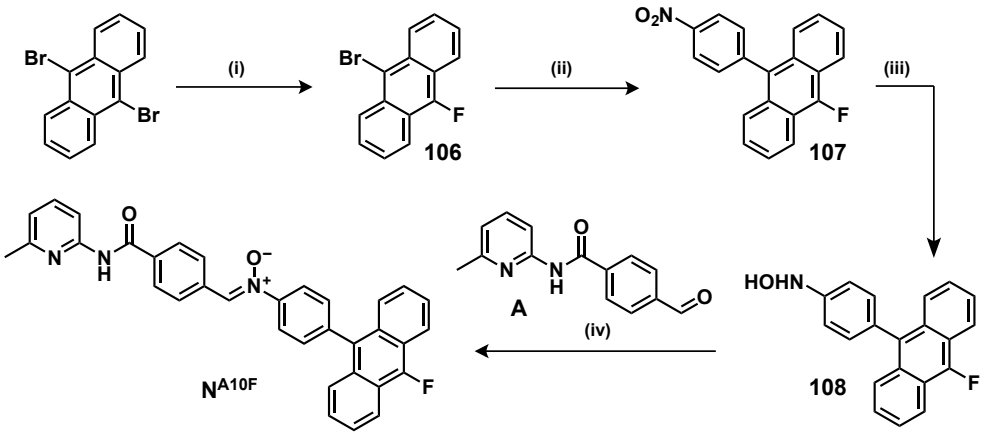
Scheme 5.4 Reagents and conditions employed in the synthesis of Nitrone $\mathbf{N^{EF}}$. (i) $Pd(PPh_3)_2Cl_2$, CuI, PPh_3, Et_3N, 90 °C, 3 days, 26%. (ii) THF, Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, RT, 25 min, quantitative conversion. (iii) EtOH/$CHCl_3$, dark, 3 days, 6 °C, 37%.

Synthesis of nitrone $\mathbf{N^{AF}}$ (**Scheme 5.5**) started by Suzuki coupling of 1-bromo-2-fluoronitrobenzene and 9-anthraceneboronic acid. Several bases and solvent mixtures were screened, however, the coupling proceeded efficiently (77%), with limited formation of anthracene-based side-product, only in the presence of K_3PO_4 as a base and $Pd(PPh_3)_4$ as a catalyst. The key intermediate nitro compound **104** was reduced to the corresponding hydroxylamine **105** using the protocol described above. The condensation of hydroxylamine **105** with aldehyde **A** in EtOH/$CHCl_3$ afforded the desired nitrone $\mathbf{N^{AF}}$ (41%) with suitable optical properties.



Scheme 5.5 Reagents and conditions employed in the synthesis of Nitrone $\mathbf{N^{AF}}$. (i) 1-bromo-2-fluoronitrobenzene, $Pd(PPh_3)_4$, K_3PO_4, PhMe/EtOH/H_2O, 100 °C, 16 h, 77%. (ii) THF, Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, RT, 35 min, quantitative conversion. (iii) EtOH/$CHCl_3$, dark, 3 days, 6 °C, 41%.

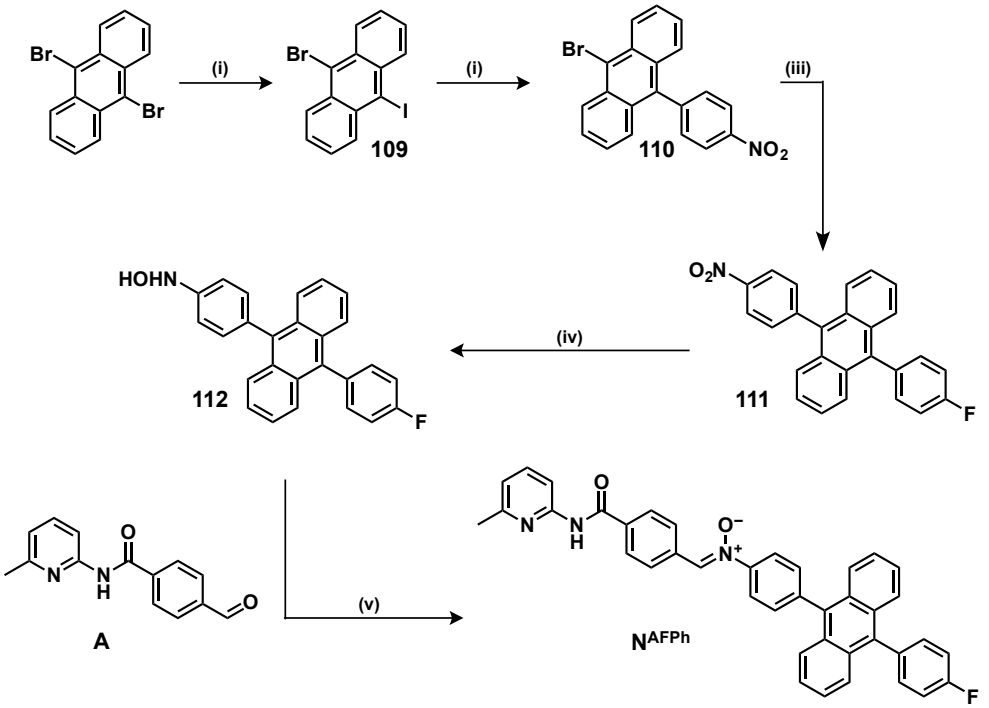
Synthetic routes to nitrones $\mathbf{N^{A10F}}$ and $\mathbf{N^{AFPh}}$ started from the same starting material, 9,10-dibromoanthracene. In order to prepare $\mathbf{N^{A10F}}$ (**Scheme 5.6**), the 9,10-dibromoanthracene was first transformed using *N*-fluorobenzenesulfonimide (NFSI) as an electrophilic source of fluorine and *n*-BuLi in Et_2O to 9-bromo-10-fluoroanthracene **106**. Despite repeated recrystallisations and thorough drying of all reagents over P_2O_5 prior to reaction, the product mixture generally contained *ca.* 20% impurities: the unreacted starting material, 9,10-difluoroanthracene and 9-bromoanthracene. The 9-bromo-10-fluoroanthracene **106** was coupled directly to 4-nitrphenylboronic acid *via* Suzuki coupling, to provide the key nitro intermediate **107** (66%). Compound **107** was again reduced to hydroxylamine **108** without any problems, using hydrazine monohydrate and a rhodium on carbon catalyst (5% wt). Condensation of **108** with aldehyde **A** in $EtOH/CHCl_3$ afforded the target nitrone $\mathbf{N^{A10F}}$ (34%).



Scheme 5.6 Reagents and conditions employed in the synthesis of Nitrone $\mathbf{N^{A10F}}$. (i) NFSI, *n*-BuLi, Et_2O, 30 °C, 16 h, *ca.* 17% (ii) 4-nitrophenylboronic acid, $Pd(PPh_3)_4$, K_3PO_4, $PhMe/EtOH/H_2O$, 100 °C, 66%. (iii) THF, Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, RT, 20 min, quantitative conversion. (iv) EtOH, $CHCl_3$, dark, 4 days, 4 °C, 34%.

In the first step in the synthesis of nitrone $\mathbf{N^{AFPh}}$, 9,10-dibromoanthracene was converted to 9-bromo-10-iodoanthracene **109** through threatment with *n*-BuLi and iodine in THF (49%). The 9-bromo-10-iodoanthracene produced was coupled selectively at the 10- position to 4-nitrophenylboronic acid (56%). The resulting 9-bromo-10-nitrorophenyl precursor **110** was coupled to 4-fluorophenylboronic acid using a second Suzuki coupling (82%). The formed fluorinated nitro compound **111** was reduced to the hydroxylamine **112** using hydrazine monohydrate in the presence of a rhodium on carbon catalyst (5% wt). In this case, hydroxylamine **112** was found to be significantly less soluble in $CDCl_3$ than the hydroxylamines synthesised previously. Nevertheless, the condensation reaction between hydroxylamine **112** and aldehyde **A** was attempted. As a result of the solubility problems, however, the reaction proceeded with very low conversion and the majority of the reaction mixture consisted of the unreacted hydroxylamine

and the corresponding azoxy side product—only a very small quantity of the desired nitrone $\mathbf{N^{AFPh}}$ was isolated by flash column chromatography of the crude mixture—a quantity sufficient only for analysis by ^{1H} NMR and $^{19}F\{^{1}H\}$ NMR spectroscopy. As a result of these difficulties in the later stages of the synthetic route to nitrone $\mathbf{N^{AFPh}}$, which were likely to translate into replicators with significant solubility problems, this optical tag was not pursued further for use in reaction-diffusion experiments.



Scheme 5.7 Reagents and conditions employed in the synthesis of Nitrone $\mathbf{N^{AFPh}}$. (i) *n*-BuLi, I_2, THF, −70 °C to RT, 49%. (ii) 4-Nitrophenylboronic acid, $Pd(OAc)_2$, Na_2CO_3, dppf, THF/H_2O, 90 °C, MW, 2.5 h, 56%. (iii) 4-Fluorophenylboronic acid, $Pd(OAc)_2$, Na_2CO_3, dppf, dioxane/H_2O, 110 °C, MW, 4.0 h, 82%. (iv) THF, Rh/C (5% wt), $NH_2NH_2{\cdot}H_2O$, RT, 30 min, quantitative conversion. (v) EtOH/$CHCl_3$, dark, 3 days, 4 °C.

5.5.1.2 NMR analysis and examination of optical properties

Prior to employing the newly designed optical nitrones, $\mathbf{N^{A}}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$ within reaction-diffusion environment, the optical signatures of these components associated with replication were examined. Very simply, a 5 mM solution of each nitrone was prepared in $CDCl_3$ (1 mL) in a glass vial and these vials were, together with a sample of nitrone $\mathbf{N^{E}}$, illuminated using a long wavelength UV light (365 nm) (**Figure 5.26a**, bottom). **Figure 5.26a** shows that all five nitrones exhibit bright yellow-green fluorescence—the observed colour of nitrones $\mathbf{N^{E}}$ and $\mathbf{N^{EF}}$ is more yellow than for the colour of the three remaining nitrones, lacking the tripe bond—an observation

that correlates with the more strongly yellow-coloured solutions that these two nitrones form in $CDCl_3$ (**Figure 5.26a**, top).

Having established that all the newly designed nitrones exhibit yellow fluorescence similar to that in nitrone $\mathbf{N^E}$, these samples were mixed with one equivalent of maleimide **M1** and left to react for *ca.* 24 hours at 5 °C—a time frame that should be sufficient for significant conversion of the nitrones into the *trans* cycloadducts. Examination of these five vials by illumination under a long wavelength UV light (365 nm) (**Figure 5.26b**) showed that the yellow fluorescence in nitrones $\mathbf{N^E}$, $\mathbf{N^A}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$ was replaced, in each case, with intense blue fluorescence, associated with the formation of products $\mathbf{T1^E}$, $\mathbf{T1^A}$, $\mathbf{T1^{EF}}$, $\mathbf{T1^{AF}}$ and $\mathbf{T1^{A10F}}$.

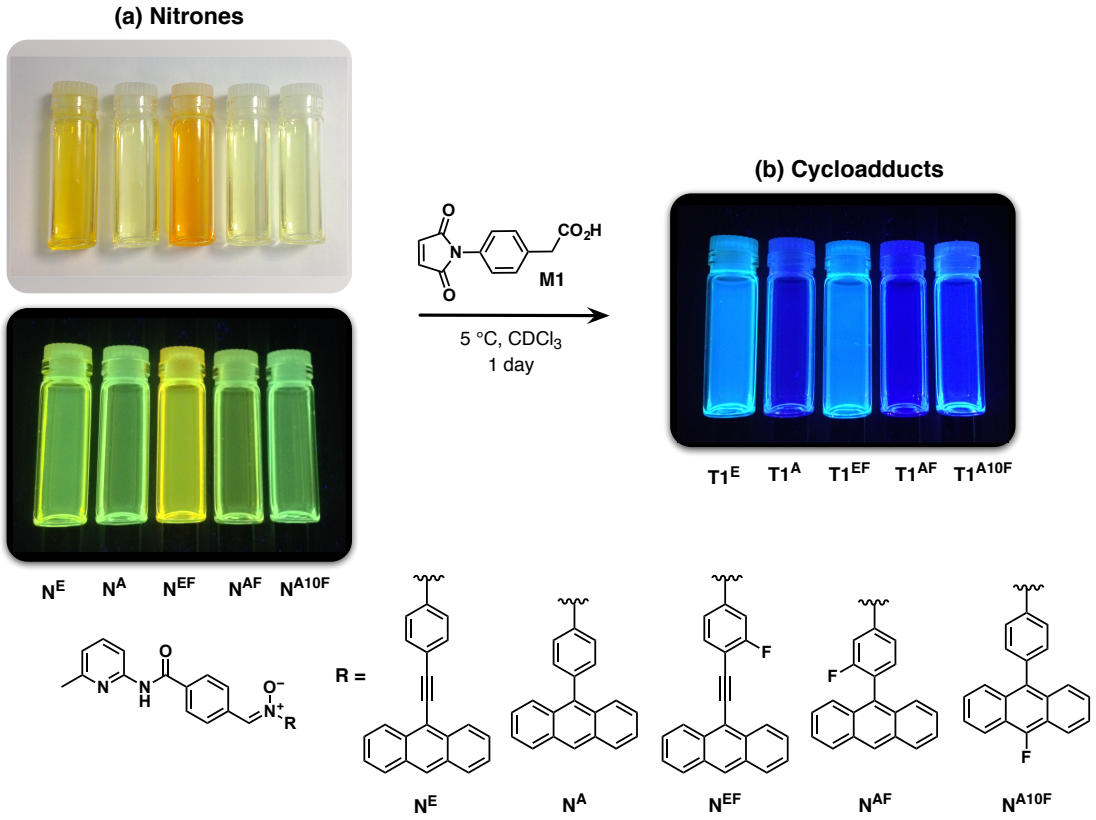


Figure 5.26 Unprocessed images of glass vials containing **(a)** solutions of nitrones (left to right) $\mathbf{N^E}$, $\mathbf{N^A}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$ at *ca.* 5 mM in $CDCl_3$, obtained by illumination with a 365 nm UV light. **(b)** Following addition of 1 equivalent of maleimide **M1** (5 mM), the fluorescent nitrones are converted to cycloadducts: $\mathbf{T1^E}$, $\mathbf{T1^A}$, $\mathbf{T1^{EF}}$, $\mathbf{T1^{AF}}$ and $\mathbf{T1^{A10F}}$, which exhibit blue fluorescence upon illumination with a 365 nm UV light. Images of blue-fluorescent cycloadducts were collected after 24 hours at 5 °C.

The number of kinetic experiments that would need to be carried out in order to comprehensively characterise the reaction of each newly designed nitrone with both maleimide **M1**, **M2** and a recognition-disabled counterpart is extremely large, requiring significant amount of NMR time. As a first step in the characterisation of these nitrones, the reaction of two nitrones, $\mathbf{N^A}$ and $\mathbf{N^{AF}}$, with maleimide **M1** was examined by NMR

spectroscopy. These two nitrones were selected because they represent model systems that allow the effect of the presence of the phenylantrhacene group and the fluorine in *meta* position on the phenyl ring to be examined. To this end, the reaction of each nitrone with **M1** was examined under experimental conditions identical to those employed in **Chapter 3** for **T1** and **T2**: 5 mM, 5 °C, $CDCl_3$. The concentration *vs* time profiles for the formation of *trans*-$\mathbf{T1^A}$ (**Figure 5.27**) was determined at periodic intervals over 16 hours by 499.9 MHz ^{1}H NMR, relative to 2,4-dinitrotoluene as internal standard. The reaction time course exhibited a typical lag period at the beginning of the reaction—after four hours, the *trans*-$\mathbf{T1^A}$ reached concentration of 0.76 mM, which further increased to 3.90 mM after 16 hours. The ratio of [*trans*]/[*cis*] diastereoisomers after 16 hours was determined as 40—a slight decrease in diastereoselectivity for the *trans* product relative to the **T1** replicator, where formation of *cis* diastereoisomer was not observed. Nevertheless, the replicator $\mathbf{T1^A}$ formed the *trans* cycloadduct very efficiently and with high diastereoselectivity—both features important for efficient reaction-diffusion front formation.

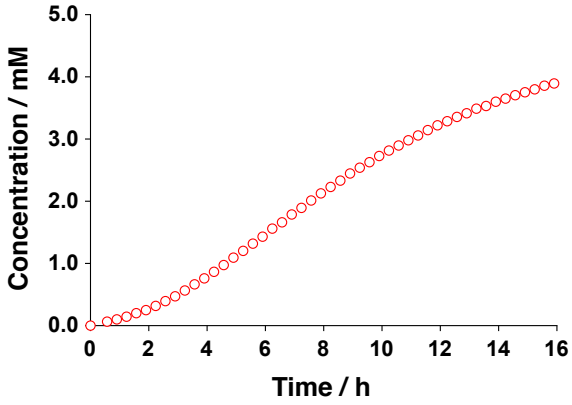


Figure 5.27 Formation of *trans*-$\mathbf{T1^A}$ (empty red circles) replicator from $\mathbf{N^A}$ and **M1** as determined by ^{1}H NMR spectroscopy: concentrations were determined relative to 2,4-dinitrotoluene as internal standard (5 °C, $CDCl_3$, 16 hours). *Cis* products are omitted for clarity.

Formation of replicator *trans*-$\mathbf{T1^{AF}}$ (**Figure 5.28a**) was examined by 470.3 MHz ^{19}F{^{1}H} NMR spectroscopy, relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard. The concentration *vs* time for this replicator, similarly to *trans*-$\mathbf{T1^A}$ and *trans*-**T1**, also showed a lag period. After four hours, the *trans* diastereoisomer reached a concentration of 0.97 mM. At the end of the reaction time (16 hours), the concentration of *trans*-$\mathbf{T1^{AF}}$ was determined as 3.90 mM.

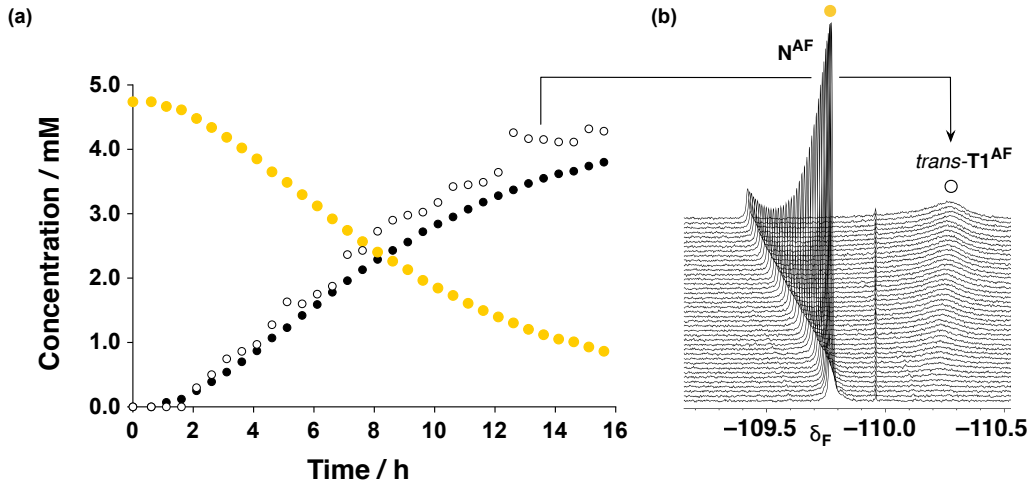


Figure 5.28 **(a)** Formation of *trans*-**T1**$^{\mathbf{AF}}$ replicator from **N**$^{\mathbf{AF}}$ and **M1** as determined by $^{19}F\{^{1}H\}$ NMR spectroscopy: concentrations were determined relative to 1-bromo-2-fluoro-4-nitrobenzene as internal standard (5 °C, $CDCl_3$, 16 hours). Empty black circles represent *trans*-**T1**$^{\mathbf{AF}}$ determined directly by deconvolution of the *trans* resonance arising from *trans*-**T1**$^{\mathbf{AF}}$ the signal and filled black circles show the expected reaction time course for *trans*-**T1**$^{\mathbf{AF}}$ calculated using nitrone consumption. *Cis* products are omitted for clarity. **(b)** Stack of of partial arrayed 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectra showing the resonances, and the progressive changes in their chemical shift, corresponding to the fluorine atom associated with nitrone **N**$^{\mathbf{AF}}$ and the *trans* diastereoisomer of **T1**$^{\mathbf{AF}}$ formed during the reaction of **N**$^{\mathbf{AF}}$ and **M1**.

The reaction time course established for replicator *trans*-**T1**$^{\mathbf{AF}}$ by analysis of the fluorine resonance associated with the *trans* cycloadduct showed a certain level of noise in the data (**Figure 5.28b**). Whilst the fluorine resonance corresponding to the nitrone **N**$^{\mathbf{AF}}$ (**Figure 5.28**, yellow) remained sharp and well-defined throughout the kinetic experiment, the resonance arising from the *trans* cycloadduct was extremely broad. The broadness associated with this *trans* peak rendered deconvolution significantly more challenging. In order to better grasp the replicator behaviour, the decrease in the nitrone resonance over time was used to calculate the amount of cycloadduct species formed in the system and the resulting concentration *vs* time plot for this conversion was compared to the data determined by analysis of the fluorine resonance corresponding to the *trans* cycloadduct (**Figure 5.28a**). The two time courses for the formation of *trans*-**T1**$^{\mathbf{AF}}$ showed very similar behaviour, but the data obtained using the changes in nitrone concentrations was significantly less noisy. The ratio of [*trans*]/[*cis*] formed in the reaction of **N**$^{\mathbf{AF}}$ and **M1** was determined from a 499.9 MHz ^{1}H NMR spectrum after 16 hours to be *ca.* 70—a slight increase in diastereoselectivity relative to the replicator **T1**$^{\mathbf{A}}$.

In order to compare the kinetic behaviour of **T1**$^{\mathbf{A}}$ and **T1**$^{\mathbf{AF}}$ with that of the original aryl-fluorine replicator **T1**, the kinetic profiles determined for these three species at identical reaction conditions were overlayed (**Figure 5.29**). Interestingly, the reaction

profile for **T1**$^{\mathbf{A}}$ mirrors the profile of **T1**$^{\mathbf{AF}}$ very closely. When these two replicators are compared to **T1**, however, a slight change in the time course is clear—**T1**$^{\mathbf{A}}$ and **T1**$^{\mathbf{AF}}$ maintain a higher concentration throughout the duration of the kinetic experiment than **T1**. The comparison of these three replicators suggests that the introduction of a phenylanthracene and its fluorinated analogue alters the replication efficiency relative to simple fluorophenyl group.

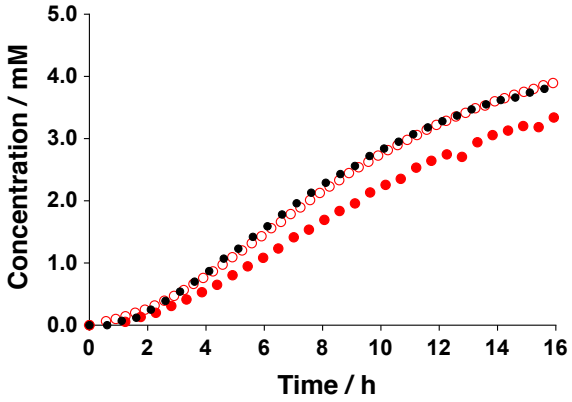


Figure 5.29 Overlay of the kinetic profiles determined for *trans*-**T1** (red circles), *trans*-**T1**$^{\mathbf{A}}$ (empty red circles) and *trans*-**T1**$^{\mathbf{AF}}$ (filled black circles) determined at 5 °C, in $CDCl_3$, over 16 hours by NMR spectroscopy.

Single crystals, suitable for X-ray crystallographic analysis, were obtained for both **T1**$^{\mathbf{A}}$ and **T1**$^{\mathbf{AF}}$ replicator: crystals of **T1**$^{\mathbf{A}}$ were grown by slow diffusion of Et_2O into a solution of **T1**$^{\mathbf{A}}$ in $CDCl_3$/PhMe, while crystals of **T1**$^{\mathbf{AF}}$ were produced through slow diffusion of Et_2O into a solution of **T1**$^{\mathbf{AF}}$ in $CHCl_3$/PhMe/THF. The X-ray analyses confirmed (**Figure 5.30**) the predicted *trans* stereochemistry of the isoxazolidine ring formed in the 1,3-dipolar cycloaddition reaction, and the 'open' conformation of the recognition sites associated with this conformation, necessary for template-directed replication. Analysis of the X-ray diffraction data showed that both **T1**$^{\mathbf{A}}$ and **T1**$^{\mathbf{AF}}$ replicators formed racemates, consisting of hydrogen-bonded dimers comprising one molecule of each enantiomer. Unexpectedly, the pyridine in the crystal structure of **T1**$^{\mathbf{AF}}$ replicator appeared to be protonated. Whilst the structure appeared to behave as the zwitterion, examination of the C–O bond length distances in the carboxylic acid group of the **M1**-arm on this replicator showed that the two C–O bonds have very different distances: (1) C–O = 1.27 Å and (2) C–O = 1.21 Å—suggesting that while the hydrogen appears to be averaged between the O and N atom in the amidopyridine unit, the pyridine is not actually protonated.

The two test NMR kinetic experiments, examining the formation of **T1**$^{\mathbf{A}}$ and **T1**$^{\mathbf{AF}}$ revealed that, whilst the initial kinetic analysis of 9-ethynylanthracene tagged replicator **T1**$^{\mathbf{E}}$, reported by Bottero and co-workers, suggested that the efficiency of replication

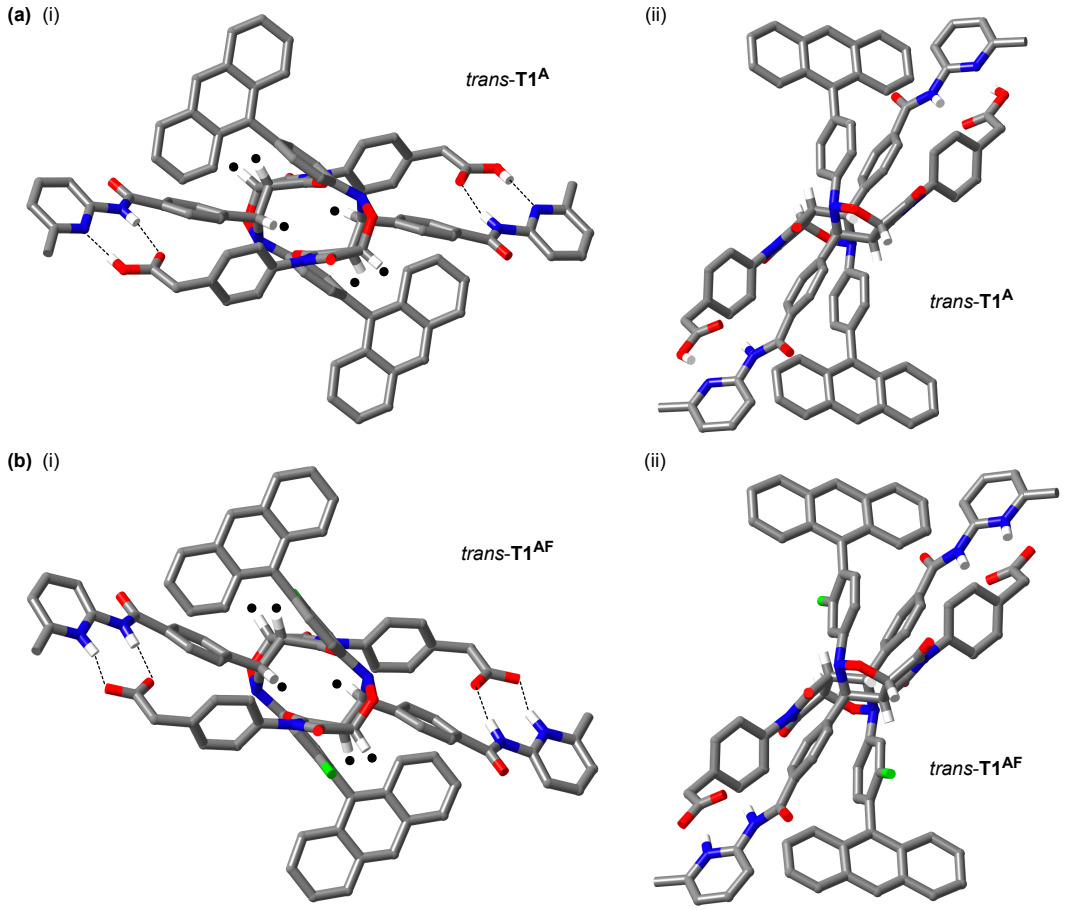


Figure 5.30 (i) And (ii) illustrate two different views of the stick representations of the X-ray crystal structures obtained for **(a) T1^{A}** and **(b) T1^{AF}** by X-ray crystallographic analysis of single crystals. C atoms are coloured grey, O atoms are in red, N atoms are in blue, H atoms are in white and F atoms are in green. Majority of hydrogen atoms are omitted for clarity. The black dots highlight the *trans* stereochemistry of the protons on the isoxazolidine ring in **T1^{A}** and **T1^{AF}**. Black dashed lines indicate interactions between the recognition sites. Data collection and X-ray diffraction analysis were performed by Dr. D. B. Cordes and Professor A. M. Z. Slawin at the University of St Andrews.

remains virtually unaffected, the introduction of an anthracene tag can alter the ability of a replicator to template its own synthesis. Despite the slight differences in reaction profiles, **T1^{A}**, **T1^{AF}** both maintained the capacity to template their own synthesis extremely efficiently and exhibited high diastereoselectivity for the *trans* cycloadduct.

5.5.1.3 Diffusion analysis

Diffusion plays a fundamental role in reaction-diffusion systems. As illustrated earlier on by the Showalter *et al*. study, the rate of diffusion can be the decisive factor in determining which autocatalyst performs the most efficiently in a system of competing replicators within the reaction-diffusion environment. Despite the vital role of diffusion in reaction dynamics, this parameter has so far remained unexamined.

Through the process of diffusion, molecules move from an area of high concentration (*i.e.* high chemical potential) to an area with low concentration, equilibrating the distribution over time and eroding any concentration gradients initially present. The term diffusion is derived from a Latin word, "diffundere" which means "to spread outwards". It should be noted that the movement of atoms and molecules through diffusion does not require any force, unlike for example the process of convection.

The reaction-diffusion experiments described in this chapter are all undertaken within the environment of a 50 µL microsyringe, with a specific internal diameter (1.03 mm). The length of the syringe corresponding to the 50 µL microsyringe is equal to 60 mm, and thus each µL is equal to 1.2 mm. The length of the syringe can be thought of as a number of connected cross-sections. Steady-state diffusion in a three-dimensional space can be described by Fick's first law of diffusion (**Equation 5.4**)—where the net flow of atoms, *i.e.* the diffusive flux (amount of material per unit of area and time)s, is equal to diffusivity times the concentration gradient. This equation can be solved for a particular diffusion coefficient *D*. *C* denotes the concentration, whereas *x* represents the position. In this law, however, diffusion is assumed to be steady-state, and, thus, constant over time.

$$J = -D\frac{\partial C}{\partial x} \tag{5.4}$$

Fick's second law of diffusion elaborates on how diffusion affects concentration change in a medium over time, as illustrated by **Equation 5.5** in one dimension.

$$\frac{\partial C_x}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{5.5}$$

Rate of diffusion in any particular system will also change with temperature (**Equation 5.6**).

$$D = D_o \times e^{\frac{-E_A}{RT}} \tag{5.6}$$

The replicators explored in this chapter, and the components necessary for their formation vary in size. The two maleimides, **M1** and **M2**, are the smallest components. and the component **M1** is slightly smaller than **M2**. The replicator templates equipped with the aryl-fluorine tags are the smallest templates in size. In contrast, replicators bearing the 9-ethynylanthracene optical tag, its fluorinated version as well as the 10-fluoroanthracene tag are amongst the largest. Diffusion coefficients of the components examined in this chapter will be bracketed by the values of the smallest and the largest components. In order to get a handle on the range of diffusion coefficients expressed in this particular system, a set of key compounds will be assessed by diffusion-ordered ^{1}H

NMR spectroscopy (DOSY). Following the initial optimisation of the DOSY dataset (in particular, the diffusion delay Δ and the diffusion gradient length δ parameters), the diffusion coefficients were determined through ^{1}H DOSY NMR for the following set of building blocks: **M1**, **M2**, $\mathbf{N^A}$ (**Table 5.2a**). Similarly, diffusion coefficients were measured also for selected templates of representative sizes, formed by the reaction of maleimdies **M1** and **M2** with three selected nitrones: $\mathbf{N^F}$, $\mathbf{N^A}$ and $\mathbf{N^{A10F}}$ (**Table 5.2b**).

Examination of the diffusion coefficients in **Table 5.2** shows that, in general, templates incorporating maleimide **M1** exhibited a slightly higher diffusion coefficient D than those formed by the reaction with maleimide **M2**. This outcome is in slight contrast to the values of D determined for **M1** and **M2** maleimides: the maleimide **M1** exhibited a marginally lower diffusion coefficient than maleimide **M2**. The diffusion coefficient of fluorescent nitrone $\mathbf{N^A}$ was found to be smaller than the D of both maleimides. Out of the six templates analysed by DOSY NMR spectroscopy, four templates: $\mathbf{T1^F}$, $\mathbf{T2^F}$, $\mathbf{T1^A}$ and $\mathbf{T1^{A10F}}$, were found to have a very similar diffusion coefficient—*ca.* $6.5 \times 10^{-10}\,\mathrm{m^2 s^{-1}}$. The two remaining templates, $\mathbf{T2^A}$ and $\mathbf{T2^{A10F}}$ exhibited a lower D—$5.35 \times 10^{-10}\,\mathrm{m^2 s^{-1}}$ and $4.86 \times 10^{-10}\,\mathrm{m^2 s^{-1}}$. These results suggest that in a competition experiment within the reaction-diffusion media, the templates formed by the reaction of **M1** can be expected to propagate at a slightly faster rate than those arising from **M2**. This difference in diffusion coefficients accounts most likely also for the slightly less efficient formation of a propagating reaction-diffusion front by $\mathbf{T2^E}$ relative to that initiated by $\mathbf{T1^E}$, discussed in the earlier sections of this chapter. Finally, this diffusion coefficient analysis suggests that the in the **T1**–**T2**-type of a network of replicators, the outcome of competition will be affected by diffusion parameter—not only is the **T1** replicator formed with a higher k_{uni} than replicator **T2**, it is also capable of diffusing through a solution of $CDCl_3$ faster, giving it an advantage in the competition for the shared nitrone building block.

Table 5.2 Diffusion coefficients (D) measured by ^{1}H DOSY NMR for a series of **(a)** building blocks necessary for the construction of two replicators, **T1** and **T2**, equipped with an example anthracene tag, and **(b)** a set of templates formed by 1,3-dipolar cycloaddition reactions between maleimdies **M1** and **M2** with a set of selected nitrones, $\mathbf{N^F}$, $\mathbf{N^A}$ and $\mathbf{N^{A10F}}$. The error in reading of the diffusion coefficients is estimated to be $\pm\ 0.25 \times 10^{-10}\ \mathrm{m^2 s^{-1}}$.

(a) Unreacted building blocks

	Diffusion coefficient (D) / $10^{-10}\ \mathrm{m^2 s^{-1}}$
$\mathbf{N^A}$	8.14
M1	11.8
M2	14.0

(b) Templates

	Diffusion coefficient (D) / $10^{-10}\ \mathrm{m^2 s^{-1}}$
T1	6.89
T2	6.15
$\mathbf{T1^A}$	6.14
$\mathbf{T2^A}$	5.35
$\mathbf{T1^{A10F}}$	6.26
$\mathbf{T2^{A10F}}$	4.86

5.5.1.4 Analysis of propagating reaction-diffusion fronts in practice: proof-of-principle experiments

A number of networks can be constructed using the four newly designed optical tags: $\mathbf{N^A}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$. The formation of some of the possible replicators has been examined through kinetic experiments, X-ray diffraction analysis and DOSY spectroscopy. In this section, a number of selected systems will be examined under competition conditions within the reaction-diffusion environment, in order to examine whether the selectivity is altered relative to that achieved within a closed reaction format. The designed reaction-diffusion experiments (**Figure 5.31**) will examine competition networks formed using two nitrones in particular: $\mathbf{N^A}$ and $\mathbf{N^{A10F}}$—these networks were selected because of their good solubility properties and the ease of their synthesis. In these reaction-diffusion experiments, the 4-fluoro-phenyl-tagged non-fluorescent templates **T1** and **T2** will be employed as the instructing templates, for the following two reasons. Firstly, a comprehensive kinetic analysis for each of the replicators incorporating the new optical tags was not undertaken, whereas replicators **T1** and **T2** have been thoroughly examined through kinetic analysis and their catalytic efficiencies and relationships have been well-established. Secondly, the reaction-diffusion experiments require the use of preformed replicator template that is different from the templates formed within the propagating front—templates **T1** and **T2** have excellent solubility properties even at concentration of 10 mM and can be detected by both ^{1}H and ^{19}F$\{^1$H$\}$

NMR spectroscopy. These two reasons make **T1** and **T2** excellent candidates for use as preformed templates in reaction-diffusion experiments—the lack of fluorescent properties should not become a problem, especially as the contents of each reaction-diffusion experiment will be divided into several sections and analysed by NMR spectroscopy.

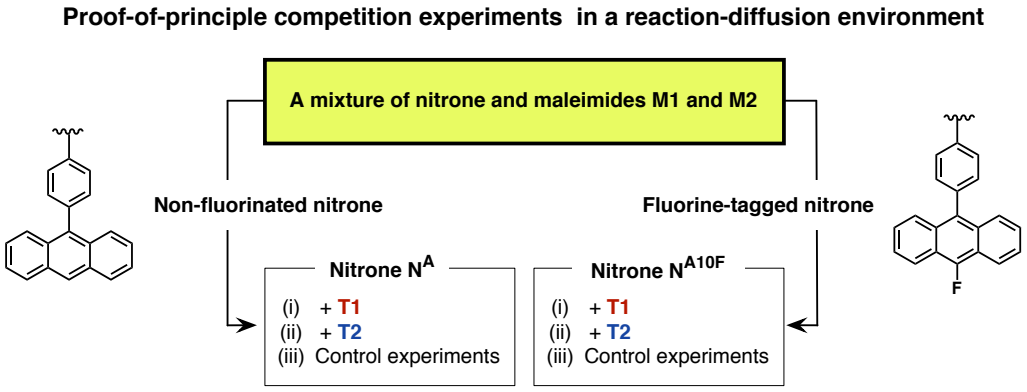


Figure 5.31 A flow chart describing the competition reaction-diffusion experiments designed to probe the selectivity in a recognition-mediated self-replicating network. A reaction mixture comprised of two maleimides, **M1** and **M2**, and either the non-fluorinated nitrone $\mathbf{N^A}$ or the fluorine-containing nitrone $\mathbf{N^{A10F}}$ will be examined. Small volume of preformed solution of (i) **T1** and (ii) **T2** will be used to initiate formation of a propagating reaction-diffusion front. (iii) The selectivity obtained within reaction-diffusion experiments will be compared to the selectivity obtained for one replicator over another within control experiments conducted in a closed reaction environment.

Combination of the two templates, **T1** and **T2**, and the two fluorescent nitrones, $\mathbf{N^A}$ and $\mathbf{N^{A10F}}$, leads to four experimental conditions overall. In each of the four experiments, the reaction conditions will be kept identical, as described in the following section and illustrated in **Figure 5.32** on an example experiment exploiting nitrone $\mathbf{N^A}$ and **T1** as template. In each experiment (**Figure 5.32a**), 3 µL of a 10 mM solution of preformed template, **T1** or **T2**[j], will be added to the syringe in order to initiate the front formation. The nitrone and maleimide reagents (47 µL) will be present at 2.5 mM—a fairly low concentration (lower than the concentration employed in **Chapter 3** and **Chapter 4**) utilised in order to reduce the rate at which the templates are formed within this mixture. Each reaction-diffusion experiment will be carried out in $CDCl_3$ and at a reaction temperature of 5 °C, parameters identical to those employed for analysis of **T1**–**T2** network previously. Each of the four described reaction-diffusion experiments will be run in seven 50 µL microsyringes in parallel, and the contents of each syringe will be divided into four sections, A to D (**Figure 5.32a**), which will be examined after six hours. Additionally, two control experiments, E and F (**Figure 5.32b**), were devised in

[j]Whilst preparation of **T1** template at a concentration of 10 mM was straightforward (no *cis* diastereoisomer is formed), formation of the catalytically-active *trans* diastereoisomer of **T2** template at the desired 10 mM concentration required that **M2** and $\mathbf{N^F}$, the components necessary for its formation, be mixed at *ca.* 12 mM.

order to examine the selectivity for one replicator over another achieved within a closed environment. Six hours after each reaction-diffusion experiment and the necessary control experiments have been started, 0.4 mL of DMSO-d_6 will be added to syringe sections A to D, and 0.3 mL of DMSO-d_6 to control experiments E and F—furnishing sufficient sample volume for analysis by NMR spectroscopy. It should be noted that throughout the analysis of reaction samples by $^{19}F\{^1H\}$ NMR spectroscopy, a delay time (d1) of 6 seconds will be used (in order to ensure complete relaxation).

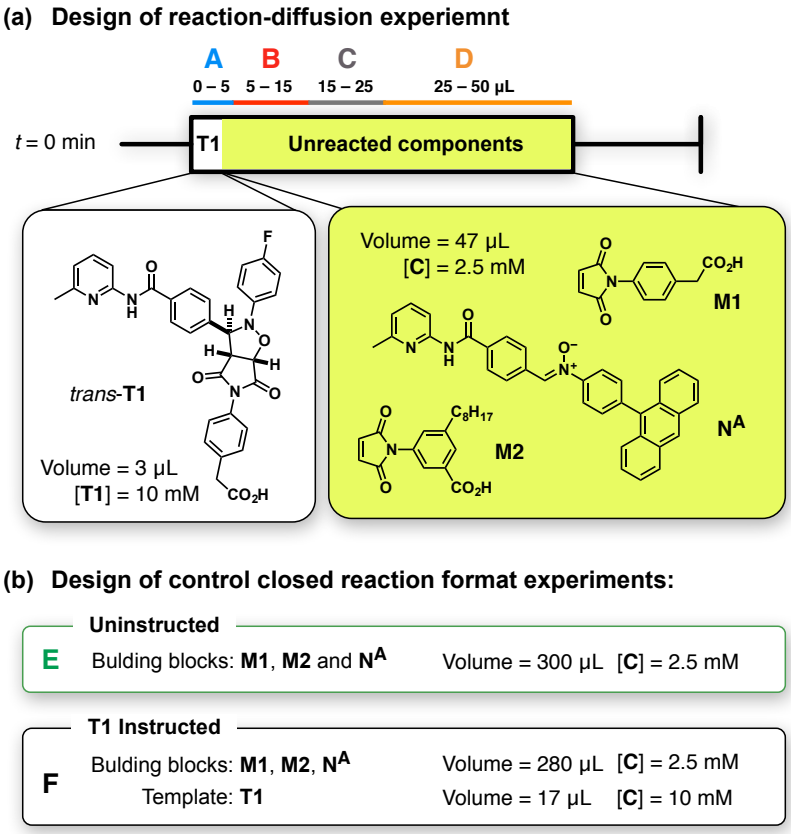


Figure 5.32 **(a)** Design of a propagating reaction-diffusion experiment, where the formation of a front is initiated by addition of 3 µL of a 10 mM solution of preformed template of *trans*-**T1** to a 2.5 mM solution of $\mathbf{N^A}$, **M1** and **M2** (47 µL). The cartoon representation of a 50 µL syringe illustrated the sections, A to D, in to which the syringe contents were divided for analysis by ^{1}H NMR spectroscopy. **(b)** Design of control experiments, E and F, undertaken to (E) examine the reaction of $\mathbf{N^{10A}}$, **M1** and **M2** (2.5 mM, 300 µL) in the absence of template and (F) the reaction of $\mathbf{N^{10A}}$, **M1** and **M2** (2.5 mM, 280 µL) in the presence of template **T1** (10 mM, 17 µL).

Using the experimental design and analytical procedure described above, the first set of experiments were carried out using the non-fluorinated nitrone $\mathbf{N^A}$, and **T1** (**Figure 5.33**) as the preformed template. The resonance associated with the *trans* protons of the isoxazolidine ring on the preformed template **T1** was detected in all conditions but E—the condition that examined the reaction of nitrone and maleimides

in the absence of template. The amount of **T1** present in the sections from the reaction-diffusion experiments relative to the formed templates progressively decreased going from section A (0–5 μL) through to D (25–50 μL), showing the extend to which this template has diffused through the syringe. It is possible, however, that the template detected in section D could have arisen as an artefact of sample manipulation, in particular the process of ejecting the syringe contents in to the NMR tube. Examination of the $_1$H NMR region typical for the *trans* resonances arising by the reaction by reaction of maleimides with nitrone $\mathbf{N^A}$ showed formation of both templates, *trans*-$\mathbf{T1^A}$ and *trans*-$\mathbf{T2^A}$, can be detected[k]. Whilst the concentration of each template formed in the various sections (A to F) cannot be determined quantitatively, the % proportion of *trans*-$\mathbf{T1^A}$ and *trans*-$\mathbf{T2^A}$ in the product pool can be determined by deconvolution of the resonances corresponding to these templates (**Figure 5.33b**)—where 100% is equal to the combined area of these two resonances.

Examination of the % of *trans*-$\mathbf{T1^A}$ and *trans*-$\mathbf{T2^A}$ in the uninstructed closed environment sample E showed that *trans*-$\mathbf{T1^A}$ is formed preferentially, at 69%, relative to *trans*-$\mathbf{T2^A}$ at 31%. The percentages of these products in condition E are identical to those determined in the syringe section D, where the added preformed template **T1** exerted minimal effect on the formation of *trans*-$\mathbf{T1^A}$ and *trans*-$\mathbf{T2^A}$. Interestingly, the proportion of *trans*-$\mathbf{T1^A}$ and *trans*-$\mathbf{T2^A}$ in D and E (*i.e.* in the absence of template) is markedly different to that determined for the **T1**–**T2** network in **Chapter 3**, where template **T2** outperformed **T1** in the uninstructed competition condition. Proportion of *trans*-$\mathbf{T1^A}$ in the **T1**-instructed control experiment increased to 77%—a percentage that is higher than that determined in the syringe sections B, C and D. Analysis of the product distribution determined in section A, however, showed that template *trans*-$\mathbf{T1^A}$ is formed at 83%—a percentage even higher than observed in the closed environment instructed experiment. Whilst the behaviour of the $\mathbf{T1^A}$–$\mathbf{T2^A}$ system seems to vary from that exhibited by the network **T1**–**T2**, the comparison of product percentages in conditions A and F suggests that the selectivity for replicator *trans*-$\mathbf{T1^A}$ over *trans*-$\mathbf{T2^A}$ is amplified by transfer of the competition to a reaction-diffusion environment.

[k]It should be noted that the chemical shift of the resonances corresponding to the *trans* protons associated with the two templates formed within each experiment: *trans*-$\mathbf{T1^A}$ and *trans*-$\mathbf{T2^A}$ as well as the preformed template added, **T1**, vary in the overlayed partial spectra shown in **Figure 5.33a**, as a result of various amounts of DMSO-d_6 added to each sample.

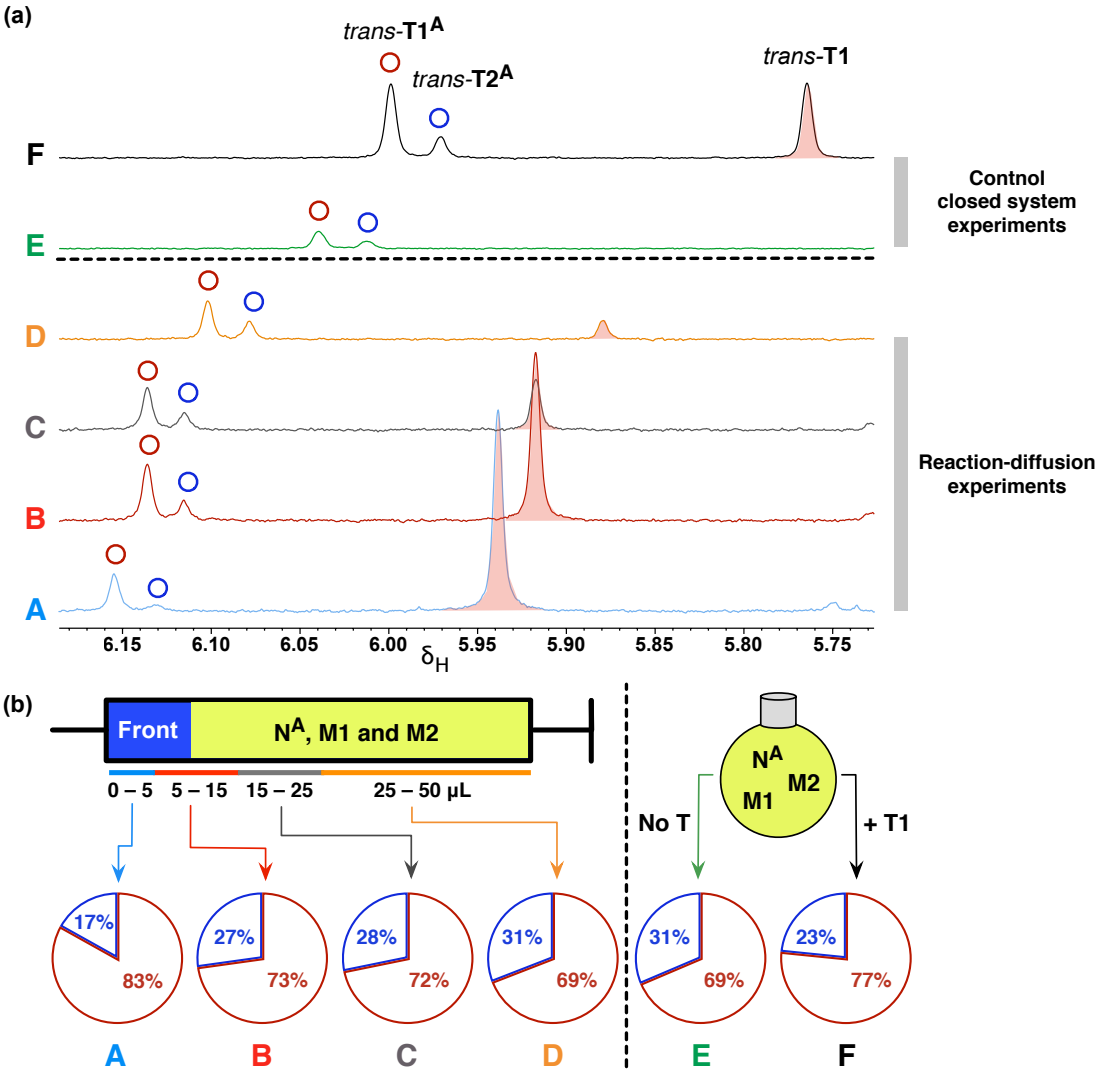


Figure 5.33 **(a)** Overlay of partial 500.1 MHz ^{1}H NMR spectra of sections A to D and control experiments E and F in an experiment examining reaction-diffusion front in a mixture of $\mathbf{N}^{\mathbf{A}}$, **M1** and **M2**, initiated by **T1** template. Each experiment was run for six hours, at which point 0.3 to 0.45 mL of DMSO-d_6 was added to each sample. All experiments were performed in $CDCl_3$ at 5 °C. In each spectrum, the centre of the DMSO-d_6 solvent resonance was used as a reference (2.50 ppm). **(b)** Pie charts showing the % of *trans*-$\mathbf{T1}^{\mathbf{A}}$ and *trans*-$\mathbf{T2}^{\mathbf{A}}$ formed within each examined section, A to F, as calculated relative to the overall combined area of the fluorine resonances arising from these two *trans* cycloadducts (100% = *trans*-$\mathbf{T1}^{\mathbf{A}}$ + *trans*-$\mathbf{T2}^{\mathbf{A}}$). *trans*-$\mathbf{T1}^{\mathbf{A}}$ and *trans*-$\mathbf{T2}^{\mathbf{A}}$ cycloadducts are shown in white and distinguished by red and blue outline, respectively.

Having established the behaviour of the replicators *trans*-**T1**$^{\mathbf{A}}$ over *trans*-**T2**$^{\mathbf{A}}$ in the presence of **T1** template, the next step was to probe the same system using **T2** as the instructing preformed template, under both reaction-diffusion and control, closed environment conditions. The overlay of partial 500.1 MHz ^{1}H NMR spectra determined for A to F (**Figure 5.34a**) showed that again, the added template **T2** can be detected in all analysed samples, but E. The percentage proportions of *trans*-**T1**$^{\mathbf{A}}$ and *trans*-**T2**$^{\mathbf{A}}$ formed in the uninstructed control experiment E (**Figure 5.34b**), 70% and 30% respectively, are essentially identical to the percentages determined for these products in the previous control experiment (**Figure 5.33b**). In the control experiment F, examining the formation of *trans*-**T1**$^{\mathbf{A}}$ and *trans*-**T2**$^{\mathbf{A}}$ in the presence of preformed **T2** in a closed reaction environment, the percentage proportion of these two templates was determined[l] as 34% and 64%, respectively—a dramatically different distribution when compared to the experiment F where the network was instructed with **T2**. The replicator *trans*-**T2**$^{\mathbf{A}}$ was formed at a higher percentage only in the section A, amounting to 73% (**Figure 5.34b**). The almost 10% increase in the proportion of *trans*-**T2**$^{\mathbf{A}}$ in condition A relative to template-instructed control experiment F suggests again, that the selectivity for one replicator over another is able to move beyond the limit imposed by kinetic selection, when examined under reaction-diffusion conditions.

Comparison of product formation in the three-building block network consisting of nitrone **N**$^{\mathbf{A}}$ and maleimides **M1** and **M2** showed that the selectivity for one replicator over another is higher under reaction-diffusion conditions, relative to the selectivity achieved in the closed reaction environment, in the presence of either **T1** or **T2** as the instructing preformed template.

[l]The partial spectrum determined for the sample F (**Figure 5.33a**) exhibits an artefact where each proton resonance is doubled, most likely as a result of problems with locking during the acquisition process, resulting from the presence of both deuterated $CDCl_3$ and DMSO-d_6—it was not possible to re-acquire the data for this sample as a result of the changes in the reaction mixture over time through the bimolecular reaction pathways (see **Chapter 9** for details).

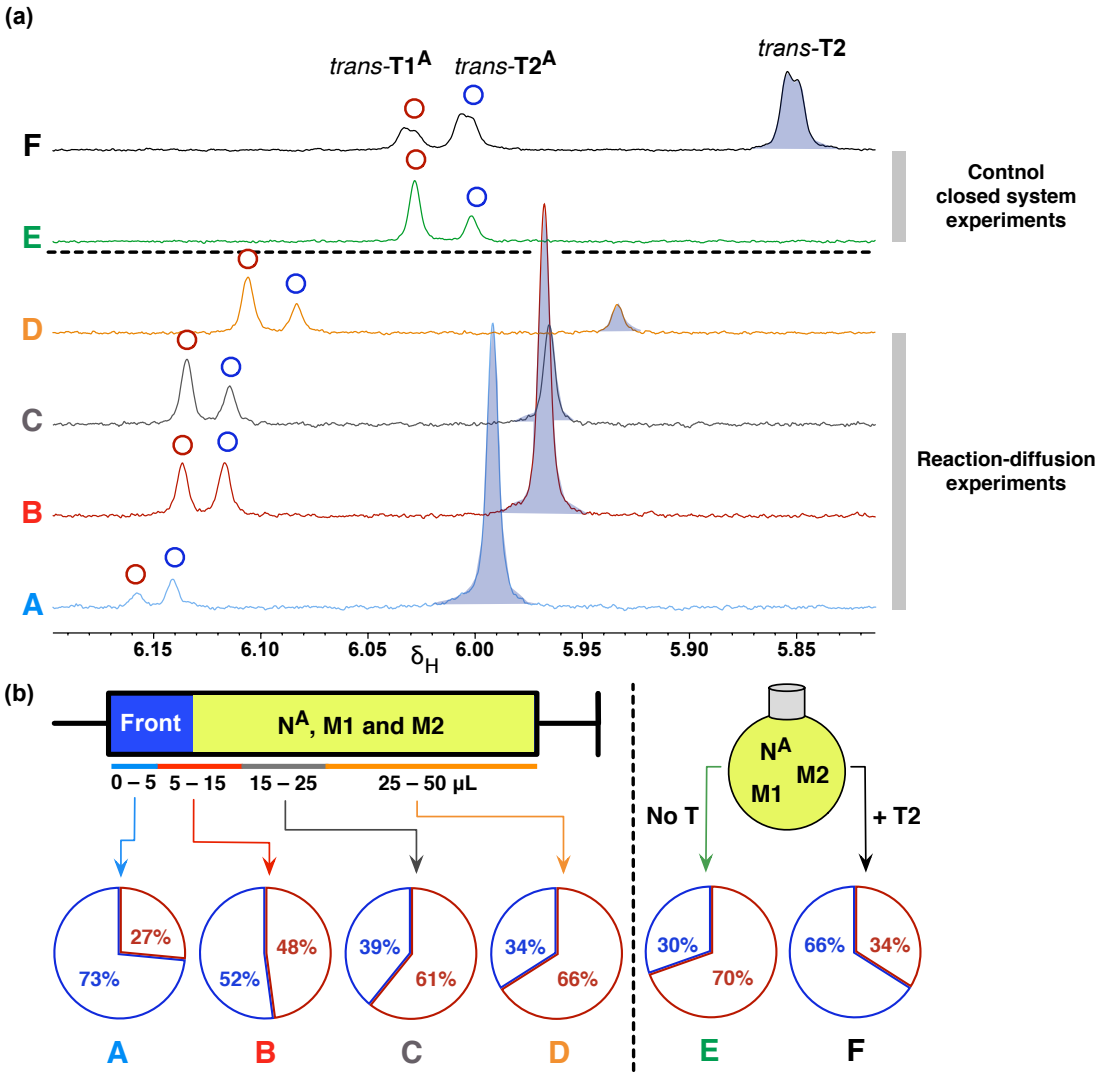


Figure 5.34 **(a)** Overlay of partial 500.1 MHz ^{1}H NMR spectra of sections A to D and control experiments E and F in an experiment examining reaction-diffusion front in a mixture of **N**$^{\mathbf{A}}$, **M1** and **M2**, initiated by **T2** template. Each experiment was run for six hours, at which point 0.3 to 0.45 mL of DMSO-d_6 was added to each sample. All experiments were performed in $CDCl_3$ at 5 °C. In each spectrum, the centre of the DMSO-d_6 solvent resonance was used as a reference (2.50 ppm). **(b)** Pie charts showing the % of *trans*-**T1**$^{\mathbf{A}}$ and *trans*-**T2**$^{\mathbf{A}}$ formed within each examined section, A to F, as calculated relative to the overall combined area of the fluorine resonances arising from these two *trans* cycloadducts (100% = *trans*-**T1**$^{\mathbf{A}}$ + *trans*-**T2**$^{\mathbf{A}}$). *trans*-**T1**$^{\mathbf{A}}$ and *trans*-**T2**$^{\mathbf{A}}$ cycloadducts are shown in white and distinguished by red and blue outline, respectively.

The second set of experiments examined the network constructed from nitrone $\mathbf{N^{A10F}}$ and **M1** and **M2**, in the presence of first **T1** (**Figure 5.35**), and later also **T2** (**Figure 5.36**). An added benefit of employing a fluorine-tagged nitrone $\mathbf{N^{A10F}}$ is that the presence of this nitrone within the reaction mixture can be easily detected using $^{19}F\{^{1}H\}$ NMR spectroscopy. **Figure 5.35a** shows the overlay of partial 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectra determined for sections A to D examining the reaction-diffusion front formed in a mixture of $\mathbf{N^{A10F}}$, **M1** and **M2**, initiated by **T1** template, and the control closed conditions experiments E and F.

Analysis of the A to F samples by 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy revealed that the preformed template **T1** can be detected is A, B, C and F, but not in samples D and E. Similarly, unreacted nitrone $\mathbf{N^{A10F}}$ was found to be present in varying amounts in all samples examined. The baseline selectivity for the *trans*-$\mathbf{T1^{A10F}}$–*trans*-$\mathbf{T2^{A10F}}$ network in the control, closed reaction environment experiment E in the absence of template showed that $\mathbf{T1^{A10F}}$ constitutes 54% of the *trans* product pool and $\mathbf{T2^{A10F}}$ 46% (**Figure 5.35b**). The percentage proportions of these products in sample E, *i.e.* in the absence of instruction, are more similar to the ratio of products determined for the **T1**–**T2** network, compared to results obtained for the system based on nitrone $\mathbf{N^{A}}$. In the presence of **T1** template, the selectivity for replicator *trans*-$\mathbf{T1^{A10F}}$ in the system relative to *trans*-$\mathbf{T2^{A10F}}$ increased—*trans*-$\mathbf{T1^{A10F}}$ amounted to 71% of the *trans* product pool. The selectivity for replicator *trans*-$\mathbf{T1^{A10F}}$ over *trans*-$\mathbf{T2^{A10F}}$ in sections C and D of the reaction-diffusion experiment was only marginally higher than the selectivity determined in sample E. The selectivity in sample B, corresponding to 5–15 μL, was even higher than the selectivity achieved in the **T1**-instructed control condition F—the replicator *trans*-$\mathbf{T1^{A10F}}$ constituted 76% of the *trans* pool. Analysis of the product distribution determined in section A of the reaction-diffusion front, the 0–5 μL fraction, showed that *trans*-$\mathbf{T1^{A10F}}$ is the only cycloadduct detected in the product pool (**Figure 5.35b**). The absence of any $\mathbf{T2^{A10F}}$ cycloadducts is a significant enhancement in selectivity for the replicator *trans*-$\mathbf{T1^{A10F}}$ relative to the selectivity in a closed environment—suggesting that complete selectivity for one replicator over another might be possible under reaction-diffusion conditions.

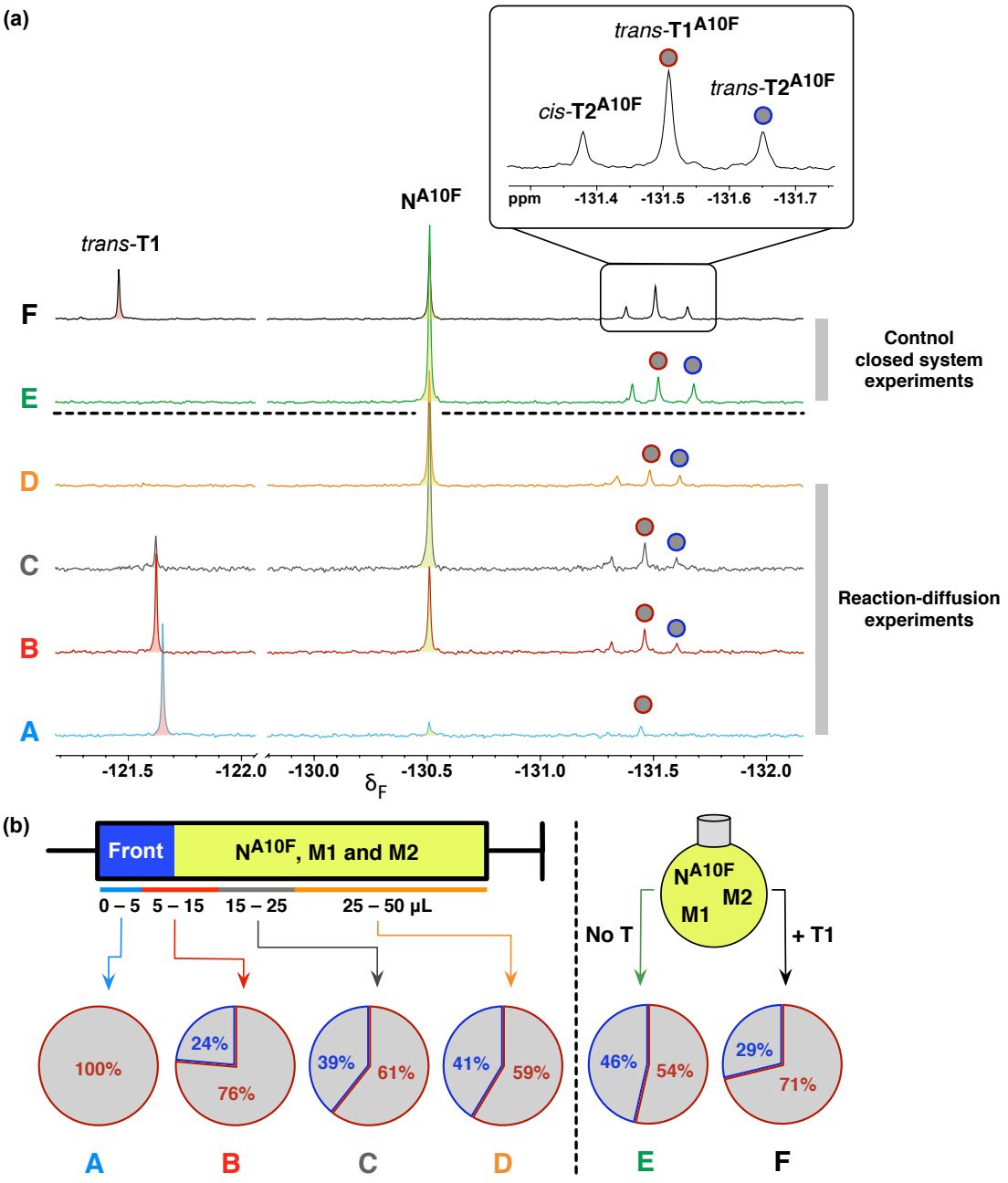


Figure 5.35 **(a)** Overlay of partial 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectra of sections A to D and control experiments E and F in an experiment examining reaction-diffusion front in a mixture of $\mathbf{N}^{\mathbf{A10F}}$, **M1** and **M2**, initiated by **T1** template. Each experiment was run for six hours, at which point 0.3 to 0.45 mL of DMSO-d_6 was added to each sample. All experiments were performed in $CDCl_3$ at 5 °C. In each spectrum, the centre of the resonance arising from nitrone $\mathbf{N}^{\mathbf{A10F}}$ was used as a reference (−130.5 ppm). **(b)** Pie charts showing the % of *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ formed within each examined section, A to F, as calculated relative to the overall combined area of the fluorine resonances arising from these two *trans* cycloadducts (100% = *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ + *trans*-$\mathbf{T2}^{\mathbf{A10F}}$). *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ cycloadducts are shown in grey and distinguished by red and blue outline, respectively.

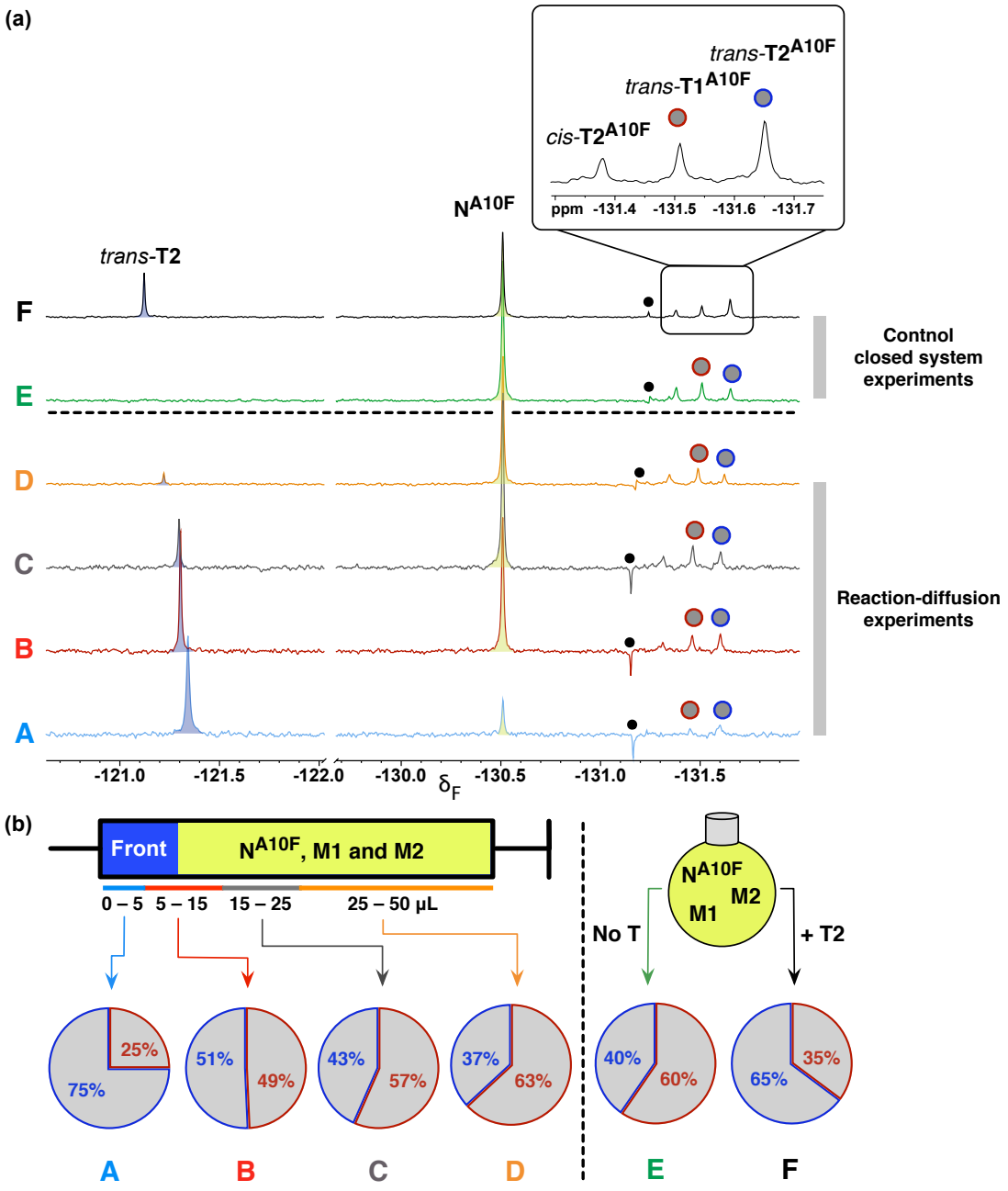


Figure 5.36 **(a)** Overlay of partial 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectra of sections A to D and control experiments E and F in an experiment examining reaction-diffusion front in a mixture of $\mathbf{N}^{\mathbf{A10F}}$, **M1** and **M2**, initiated by **T2** template. Each experiment was run for six hours, at which point 0.3 to 0.45 mL of DMSO-d_6 was added to each sample. All experiments were performed in $CDCl_3$ at 5 °C. In each spectrum, the centre of the resonance arising from nitrone $\mathbf{N}^{\mathbf{A10F}}$ was used as a reference (−130.5 ppm). **(b)** Pie charts showing the % of *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ formed within each examined section, A to F, as calculated relative to the overall combined area of the fluorine resonances arising from these two *trans* cycloadducts (100% = *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ + *trans*-$\mathbf{T2}^{\mathbf{A10F}}$). *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ cycloadducts are shown in grey and distinguished by red and blue outline, respectively. Black circle represents a resonance in samples A to F, the identity of which could not be established.

In the last experiment, the selectivity in the network constructed from nitrone $\mathbf{N^{A10F}}$ and **M1** and **M2** was analysed in the presence of preformed template **T2** (**Figure 5.36**). **Figure 5.36a** shows the overlay of partial 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectra determined for sections A to D examining the reaction-diffusion front formed in a mixture of $\mathbf{N^{A10F}}$, **M1** and **M2**, initiated by **T2** template, and the control closed conditions experiments E and F, and **Figure 5.36a** portrays the proportion of *trans*-$\mathbf{T1^{A10F}}$ and *trans*-$\mathbf{T2^{A10F}}$ in the *trans* product pool. The selectivity in the $\mathbf{N^{A10F}}$-based network in a closed environment, in the absence of template (condition E) revealed that products *trans*-$\mathbf{T1^{A10F}}$ and *trans*-$\mathbf{T2^{A10F}}$ account for 60% and 40%, respectively—percentage values that are fairly similar to those determined for sample E in **Figure 5.35b**. When the reaction mixture was instructed with preformed **T2**, the system exhibited an increased preference for *trans*-$\mathbf{T2^{A10F}}$, which constituted 65% of the *trans* product pool (*trans*-$\mathbf{T1^{A10F}}$ = 35%). The selectivity for *trans*-$\mathbf{T2^{A10F}}$ over *trans*-$\mathbf{T1^{A10F}}$ was lower in sections B, C and D of the reaction-diffusion experiment, initiated by **T2**. Only section A, examining the 0–5 μL fraction, showed higher selectivity for the replicator $\mathbf{T2^{A10F}}$ than the selectivity achieved in the template-instructed analogous experiment in a closed system (condition F).

Finally, the comparison of the selectivities achieved in the control, closed system experiment F, in which preformed template is added, to the selectivities determined in section A (0 to 5 μL) of each reaction-diffusion experiment across the four sets of experimental conditions revealed that the reaction-diffusion environment results, in each case, in a higher selectivity for one replicator over another—a selectivity beyond that achievable in a network of replicators with a shared building blocks through kinetic selection only.

Examination of the selectivity for one replicator over another in the two reaction networks, based on nitrone $\mathbf{N^{A}}$ and nitrone $\mathbf{N^{A10F}}$, showed that their behaviour is slightly different from each other, as well as from the behaviour of the **T1**–**T2** network. Both networks, exploiting anthracene-based optical tags exhibited a stronger preference for the formation of the replicator formed by reactions with maleimide **M1** as opposed to **M2**—an outcome in contrast to that of the **T1**–**T2** network where the concentration of **T2** replicator exceeded the quantity of **T1** in the uninstructed competition experiment within a closed reaction environment (**Chapter 3**). In order to further examine the levels of selectivity achieved in the *trans*-$\mathbf{T1^{A}}$–*trans*-$\mathbf{T2^{A}}$ and *trans*-$\mathbf{T1^{A10F}}$–*trans*-$\mathbf{T2^{A10F}}$ networks of replicators, the product ratios determined in the two **T1**-instructed (**Figure 5.37a**) and two **T2**-instructed (**Figure 5.37b**) experiments were compared.

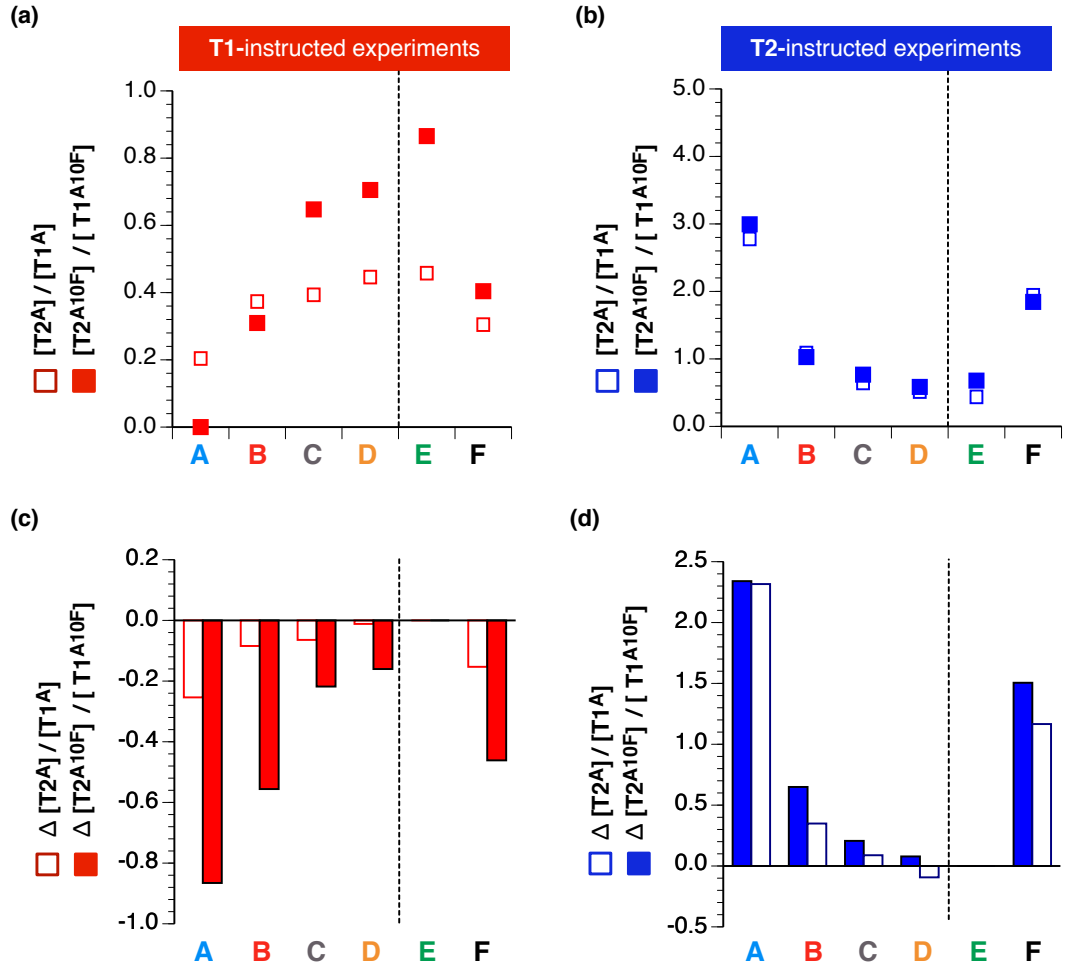


Figure 5.37 Comparison of *trans* product ratios determined in samples A to F across the sets of **(a)** **T1**-instructed and **(b)** **T2**-instructed experiments. **(a)** **T1**-instructed experiments show the ratio of *trans*-$\mathbf{T2}^{\mathbf{A}}$ over *trans*-$\mathbf{T1}^{\mathbf{A}}$ (white squares with a red line) and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ over *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ (red squares). **(b)** **T2**-instructed experiments examined the selectivity for *trans*-$\mathbf{T2}^{\mathbf{A}}$ over *trans*-$\mathbf{T1}^{\mathbf{A}}$ (white squares with a blue line) and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ over *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ (blue squares). Change in product ratios determined in samples A to F across the sets of **(c)** **T1**-instructed and **(d)** **T2**-instructed experiments. **(c)** Change in the ratio of *trans*-$\mathbf{T2}^{\mathbf{A}}$ over *trans*-$\mathbf{T1}^{\mathbf{A}}$ (white squares with a red line) and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ over *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ (red squares), calculated relative to the product ratio in condition E. **(d)** Change in the ratio of *trans*-$\mathbf{T2}^{\mathbf{A}}$ over *trans*-$\mathbf{T1}^{\mathbf{A}}$ (white squares with a blue line) and *trans*-$\mathbf{T2}^{\mathbf{A10F}}$ over *trans*-$\mathbf{T1}^{\mathbf{A10F}}$ (blue squares), calculated relative to the product ratio in condition E.

In line with the [**T2**]/[**T1**] ratios examined in **Chapter 3** and **Chapter 4**, the **T1**-instructed experiments (and the associated closed system experiments) examined the effect of varying the reaction conditions on the ratio of the product incorporating maleimide **M2** (*trans*-$\mathbf{T2^{A}}$ or *trans*-$\mathbf{T2^{A10F}}$) over the replicator formed by the reaction of nitrone with **M1** (*trans*-$\mathbf{T1^{A}}$ or *trans*-$\mathbf{T1^{A10F}}$). Similarly, the selectivity for the **M2**-based replicators over those formed by the reaction of nitrone with **M1** was examined in the **T2**-instructed set of experiments.

Comparison of the selectivities achieved for one replicator over another in the two **T1**-instructed experiments (**Figure 5.37a**) showed a number of trends. Generally, the selectivity for one replicator over another is generally very similar in sections C and D of the syringe (15 to 50 µL) and the control uninstructed condition E—suggesting that front-propagation does not extend beyond the 15 µL in any of the examined reaction-diffusion experiments. Instead, the nitrone and maleimide components in sections C and D react in an environment that is essentially undistinguishable from that of the control uninstructed experiment (E). Interestingly, the product ratios determined in the two sets of **T2**-instructed experiments were extremely similar. The same trend was not observed for the **T1**-instructed experiments—in this case, the reaction network based on nitrone $\mathbf{N^{A10F}}$ exhibited a more marked increase in the selectivity for the *trans*-$\mathbf{T1^{A10F}}$ product, compared to the $\mathbf{N^{A}}$-based system, upon instruction with preformed **T1** template. The variation in selectivity for one product over another across the six examined samples were visualised by plotting the difference between the product ratio determined in the control uninstructed experiment E and samples A to D, and F. (**Figure 5.37c** and **Figure 5.37d**). Finally, the comparison of the selectivities achieved in each template-instructed control condition F to the selectivities determined in section A (0 to 5 µL) of each reaction-diffusion experiment across the four sets of experimental conditions revealed that the reaction-diffusion environment results, in each case, in a higher selectivity for one replicator over another—a selectivity beyond that achievable through kinetic selection only in a network of replicators with a shared building block.

5.6 Summary

The work presented in this chapter described the experimental efforts undertaken to employ reaction-diffusion conditions as the means to overcoming the limits on selectivity imposed in networks of replicators in a closed reaction environment. In order to permit monitoring of the reaction progress within the environment of a 50 µL gas-tight syringe as a model of far-from-equilibrium conditions, the aryl-fluorine tag in nitrone $\mathbf{N^{F}}$ (**Figure 5.38a**), employed in **Chapter 3** and **Chapter 4**, was substituted

with an 9-ethynylanthracene tag. The reaction of this anthracene tag bearing nitrone $\mathbf{N^E}$ with a maleimide is associated with a dramatic change in fluorescence, from yellow to blue (**Figure 5.38a**).

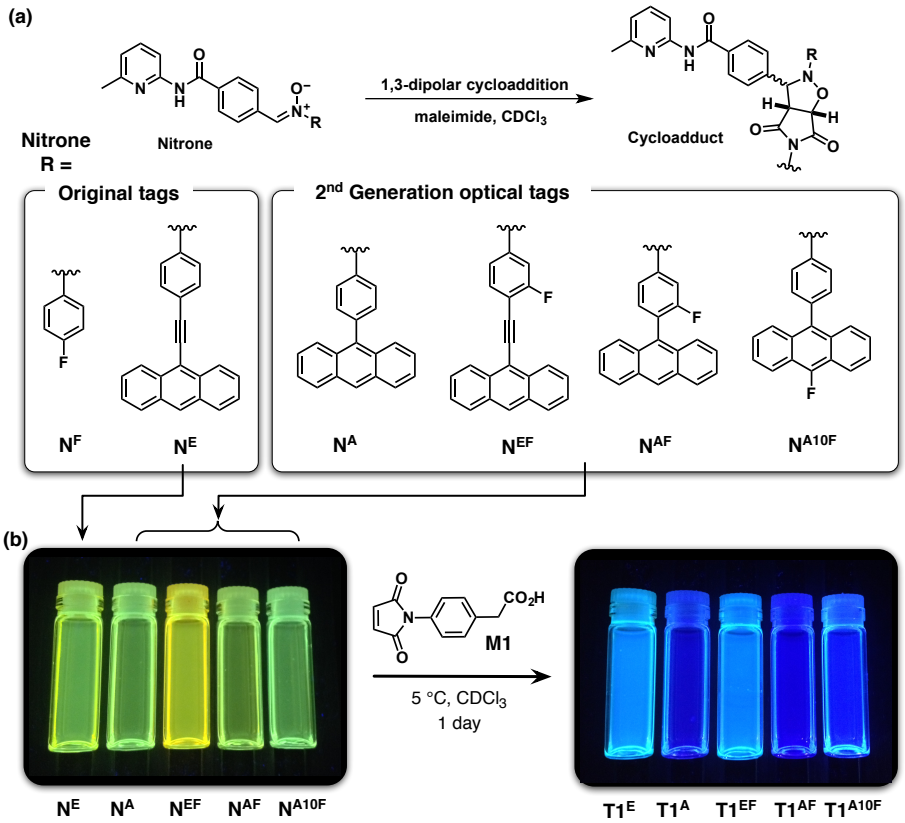


Figure 5.38 **(a)** A nitrone reacts with maleimide through 1,3-dipolar cycloaddition to afford a self-replicating template. R corresponds to the structure of the tag present in the nitrone and the replicator template. Nitrone $\mathbf{N^F}$ was used to construct the **T1**–**T2** network explored in **Chapter 3** and **Chapter 4**. This aryl-fluorine nitrone was redesigned to possess an anthracene tag, furnishing a fluorescent nitrone $\mathbf{N^E}$, that permitted permitted the replication progress to be monitored in real time under illumination by UV. Nitrones $\mathbf{N^A}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$ employ newly designed anthracene optical tags for use in the reaction-diffusion experiments. **(b)** The yellow fluorescence of nitrones $\mathbf{N^E}$, $\mathbf{N^A}$, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$ changes dramatically to blue, upon their reaction with maleimide **M1**, to produce replicators $\mathbf{T1^E}$, $\mathbf{T1^A}$, $\mathbf{T1^{EF}}$, $\mathbf{T1^{AF}}$ and $\mathbf{T1^{A10F}}$. Reaction conditions: 24 hours at 5 °C, concentration *ca.* 5 mM, $CDCl_3$, irradiation with a 365 nm UV light.

Reaction of nitrone $\mathbf{N^E}$ with maleimides **M1** and **M2** afforded a reaction network $\mathbf{T1^E}$–$\mathbf{T2^E}$, with optical properties that allowed its behaviour to be studied under reaction-diffusion conditions. Initially, the ability of each of these replicators to initiate the formation of a propagating reaction-diffusion front in response to seeding with preformed template was studied and confirmed in isolation. The network of two replicators was also studied under competition conditions, where maleimides **M1** and **M2** competed for the nitrone component $\mathbf{N^E}$. Analysis of the product distribution determined within the reaction-diffusion environment revealed an increase in selectivity for the instructed

replicator relative to the selectivity achieved in an analogous uninstructed experiment under closed reaction conditions.

Several requirements were identified for the analysis of networks of competing replicators under reaction-diffusion environment. Namely, it must be possible to distinguish the preformed template added to initiate front formation from the templates formed in the reaction—a requirement that was fulfilled by utilising preformed template bearing a different tag than the group on the nitrone element. Furthermore, a method must be devised for the analysis of the very small volumes of reaction material available from reaction-diffusion experiments—running several identical reaction-diffusion experiments in parallel was found to provide sufficient quantities of material for analysis by NMR spectroscopy. Having identified methods of satisfying these requirements, the work in this chapter focused on developing additional fluorescent tags with suitable optical properties, in particular those equipped with a fluorine tag, that would allow networks of competing replicators to be examined under reaction-diffusion conditions. Several compounds were identified as targets and successfully synthesised—a non-fluorinated anthracene tag, $\mathbf{N^A}$, and three fluorine-containing tags, $\mathbf{N^{EF}}$, $\mathbf{N^{AF}}$ and $\mathbf{N^{A10F}}$ (**Figure 5.38a**). Reaction of two nitrones, $\mathbf{N^A}$ and $\mathbf{N^{AF}}$, with maleimide **M1** was examined through kinetic experiments, showing that whilst the concentration *vs* time profiles for these replicators do not match the time-course of replicator **T1** (formed by reaction of **M1** and $\mathbf{N^F}$) exactly, both $\mathbf{T1^A}$ and $\mathbf{T1^{AF}}$ were formed very efficiently and with high diastereoselectivity for the *trans* cycloadduct. Selected nitrone, maleimide and replicator components were also analysed by DOSY NMR spectroscopy, and their diffusion coefficients D established: the results of this analysis showed that the two maleimides **M1** and **M2** have the highest diffusion coefficients, whereas the cycloadduct products formed by the reaction with **M2** possess the lowest diffusion coefficients.

Following the preliminary kinetic and diffusion analysis of the selected newly designed fluorescent tags and replicators containing them, two nitrones, $\mathbf{N^A}$ and $\mathbf{N^{A10F}}$, were employed in a series of proof-of-principle competition reaction-diffusion experiments examining the selectivity for one replicator over another, in the presence of instructing template. In each reaction-diffusion experiment, the contents of the microsyringes were divided into four sections (**Figure 5.39b**), combined with the same sections from other syringes, and the changes in product distribution within each sample examined. Using this procedure, it was possible to demonstrate how the selectivity for one replicator over another changes within the reaction-diffusion front, relative to the environment within the syringe where the front did not extend to—an environment essentially identical to that employed in the control closed reaction format.

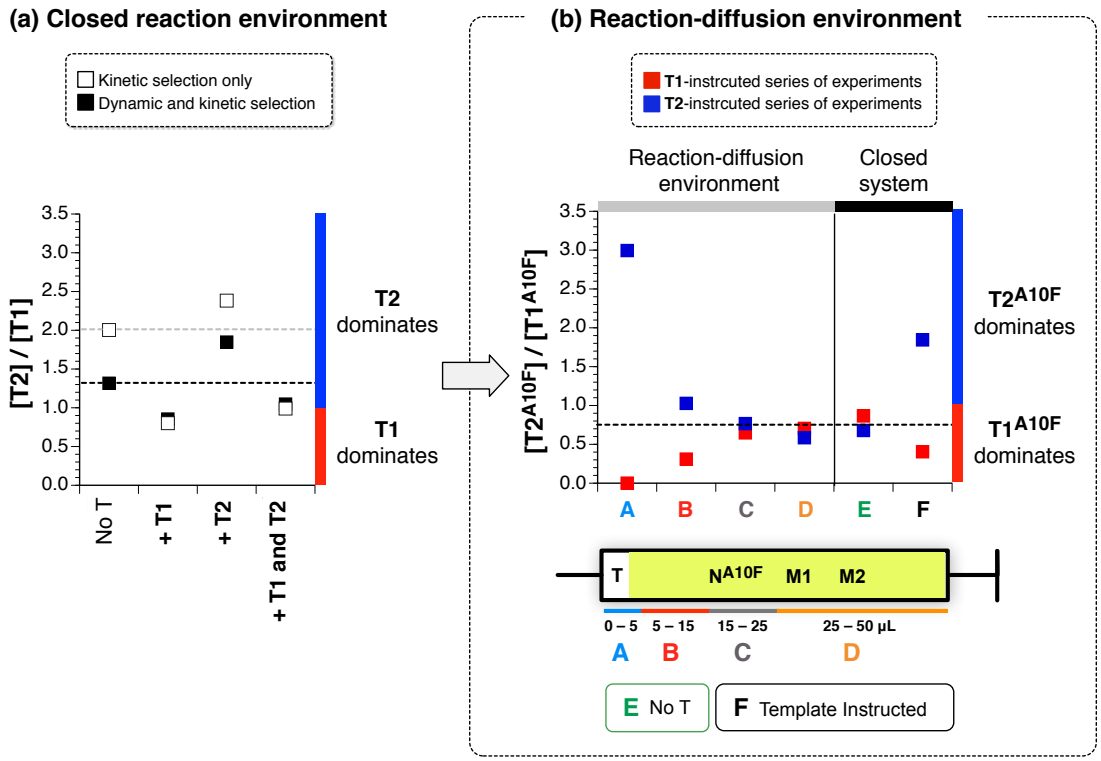


Figure 5.39 Comparison of selectivities achieved for one replicator over another **(a)** within closed reaction environment and **(b)** within a series of reaction-diffusion experiments. **(a)** The ratio of [**T2**]/[**T1**] formed in the environment of a DCL after two days (black squares) and after four hours through kinetic selection only (white squares) across four template conditions: uninstructed, **T1**, **T2** and **T1** and **T2** simultaneously, examined in **Chapter 3** and **Chapter 4**, respectively. The red and blue regions highlight the conditions where **T1** and **T2** replicators dominate, respectively. **(b)** The ratio of [$\mathbf{T2}^{\mathbf{A10F}}$]/[$\mathbf{T1}^{\mathbf{A10F}}$] determined in samples A to F across the set of **T1**- (red squares) and **T2**-instructed experiments (blue squares). Dashed lines highlight the ratios determined in each condition in the absence of template.

The highest selectivity for one replicator over another was observed in the fluorescent reaction network based on nitrone $\mathbf{N}^{\mathbf{A10F}}$ (**Figure 5.39b**)—in both the **T1**- and **T2**-instructed series of experiment, the selectivity for one product over another was significantly higher within the reaction-diffusion front compared to the instructed control experiment undertaken under closed reaction conditions. While not directly comparable, the selectivity demonstrated for one product over another within the reaction-diffusion format were higher than those achieved in the **T1**–**T2** network (**Figure 5.39a**) through kinetic selection (**Chapter 3**) and within the environment of a dynamic covalent library (**Chapter 4**).

In summary, throughout the analysis of networks of interconnected replicators under reaction-diffusion conditions, several design and analytical challenges were encountered, and some of these have been resolved successfully:

- ✓ The quantity of material available for analysis in reaction-diffusion experiments can be increased successfully by performing numerous identical experiments in parallel.

- ✓ It is possible to divide the contents of each microsyringe into several sections, which are combined from the various syringes employed and analysed individually, allowing the front progression to be determined analytically.

- ✓ The selectivity for one replicator over another achieved within the various sections of a reaction-diffusion experiment can be compared to the selectivity determined in analogous control closed environment experiments—providing a method of assessing which sections of the microsyringe exhibit increased selectivity.

- ✓ Utilising preformed template equipped with a tag that is different from the tag on the nitrone components allows the template formed within the experiment to be distinguished unambiguously from the template added to initiate the front formation.

- ✓ Addition of DMSO-d_6 to all reaction-diffusion and closed system experiments stops the recognition-mediated processes from operating in each experiment at the same time—thereby slowing down the reaction processes to the rates of the bimolecular reactions—the rate of which is further reduced by the dilution of each sample with DMSO-d_6.

Nevertheless, despite the marked progress achieved in the analysis of reaction-diffusion experiments and the selectivities for one replicator over another achieved within them, a few issues remain to be addressed:

- At this stage, it is not possible to establish if and how the manual manipulation of the microsyringes affects the outcome of the reaction-diffusion experiments.

- Comprehensive set of kinetic analyses examining the reactions between nitrones bearing the various developed optical tags and both maleimides needs to be undertaken in order to establish the efficiency of all auto- and crosscatalytic pathways.

- Kinetic fitting of the experimental data will provide access to reaction and association parameters which can be used, together with diffusion coefficients established for the reactants and template, in computational modelling studies directed at probing the behaviour of reaction-diffusion fronts driven by small-molecule based synthetic replicators.

In conclusion, the work presented in this chapter demonstrated a successful implementation of propagating reaction-diffusion fronts driven by replication processes. The reaction-diffusion environment allowed networks of interconnected replicators to be examined under conditions that allowed the selectivity for one replicator over another to move beyond the limit imposed by closed reaction conditions, *i.e*. kinetic selection—exhibiting signs of complete selectivity.

CHAPTER 6

INTEGRATING SELF-REPLICATION WITH THE FORMATION OF A [2]ROTAXANE

Some of the work presented in this chapter has been published in:

- T. Kosikova, N. I. Hassan, D. B. Cordes, A. M. Z. Slawin and D. Philp, *J. Am. Chem. Soc.* **2015**, *137*, 16074–16083
- A. Vidonne, T. Kosikova and D. Philp, *Chem. Sci.* **2016**, *7*, 2592–2603

6.1 Preamble

Molecular recognition plays a pivotal role in the self-assembly and self-organisation of biological and chemical systems alike. The recognition and reaction processes in complex biological networks, however, do not operate in isolation, often encompassing multiple levels of instruction—an area virtually unexplored in the field of chemistry. Studying the connectivity and interplay between the various components in chemical reaction networks operating more than a single recognition-mediated process simultaneously is an essential next step in determining the fundamental principles governing function expressed by both chemical and non-chemical complex systems.

The experimental chapters presented in this thesis have, thus far, examined a chemical network with the capacity to exhibit the complex phenomena of self-replication. Namely, the work focused on investigating the behaviour of an interconnected network of two structurally-similar replicators, **T1** and **T2** and the various modes of resolving this network. Specifically, the network resolution in **Chapter 3** was driven by kinetic selection, in a closed reaction environment. **Chapter 4** employed the environment

of a dynamic covalent library, where covalent and non-covalent selection processes governed simultaneously the network resolution. Finally, **Chapter 5** explored a change in the reaction format, examining the network of replicators under reaction-diffusion environment. The results of kinetic analyses confirmed that even small changes in the identity or orientation of recognition-enabling structural elements can exert profound effects on the recognition and reaction processes driving the formation of the replicators within the network. Therefore, in order to exploit the potential afforded by molecular recognition and self-assembly as the driving forces for the design and synthesis of complex systems in a pre-organised manner, consideration has to be given to both (i) the desirable function-encoding elements and (ii) their relative position within a particular system.

This chapter aims to examine a synthetic model of a multi-level instruction system by incorporating replication with another recognition-mediated process—in particular, a process leading to the formation of a mechanically-interlocked architecture—a [2]rotaxane. This chapter will provide a short introduction to the field of mechanically-interlocked architectures, with particular focus on rotaxanes, their properties, synthesis and potential applications. Subsequently, an overview of the requirements for the formation of a self-replicating rotaxane and the various possible models for the integration of self-replication with rotaxane formation will be described and put in the context of the experimental work undertaken to date.

6.2 Mechanically-interlocked architectures

In addition to formation of covalent bonds, two or more molecules can also interact and associate together *via* molecular recognition, *i.e*. through non-covalent interactions, to form self-assembled systems with an organisational hierarchy. In addition to the more traditional type of non-covalent interactions, molecules can be held together by the formation of a mechanical[243,244] (also referred to as topological[245]) bond—a bond that links two or more molecules in such a manner that their separation would necessitate the breaking of at least one covalent bond. In such a case, the resulting mechanically-interlocked architecture (MIA) is considered to be a single molecule rather than a supramolecular assembly comprised of individual components. Formation of such architectures exploits tools from both supramolecular chemistry and traditional synthetic chemistry, producing molecules with varied topology, *e.g*. catenanes,[246–248] rotaxanes[249] and molecular knots[250–253] (**Figure 6.1**), to name a few.

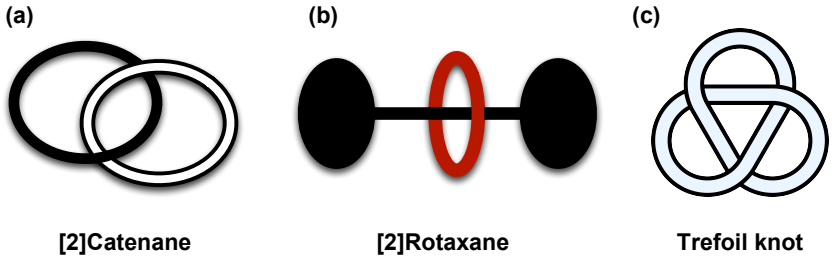


Figure 6.1 Cartoon representation of several classes of mechanically-interlocked architectures. **(a)** [2]Catenane comprised of two interlocked macrocycles; **(b)** [2]Rotaxane architecture built from a macrocycle component encircling a thread equipped with bulky stoppers and **(c)** trefoil molecular knot—an architecture similar to that of a macroscopic knot.

6.2.1 Chemical topology, nomenclature and synthetic approaches

The term *chemical topology* was first introduced[245] in 1961 by Frisch and Wasserman in a seminal work bearing the same title. Chemical topology provides a way of distinguishing between several topological isomers, *i.e*. species that cannot be sufficiently distinguished using the number and order of atoms within it, the bonds connecting them and their relative spatial arrangement. The concept of chemical topology can be demonstrated using one of the examples provided[245] by Frisch and Wasserman—cycloalkanes longer than 20 carbon atoms can exist as either individual rings (**Figure 6.2a**) or if the cavity is sufficiently large, as systems of interlocked rings, the so called catenanes (**Figure 6.2b**). These two forms are topological isomers of each other, with potentially very different properties, and it is not possible to simply convert one form in to the other. The authors highlight that, while the rings in the [2]catenane are connected through a topological bond, the strength (*i.e*. the stability) of this MIA as a whole remains the same as the strength the non-interlocked counterpart—*i.e*. only as strong as the covalent bonds within them. Interestingly, a rotaxane molecular architecture, which is assembled from a macrocycle that encircles a threading component equipped with bulky stoppering groups at each end, is not a topological isomer of its parts, because the components can be separated without the need to break any covalent bonds, for example, simply by deformation of the macrocycle.

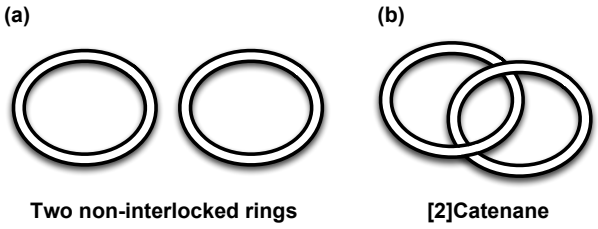
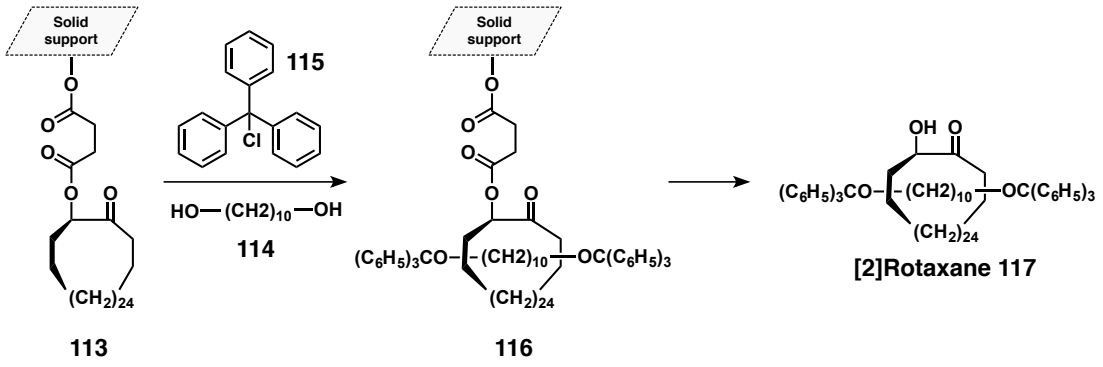


Figure 6.2 Long carbon chain cycloalkanes can exist as two topological isomers: **(a)** two separate non-interlocked macrocycles and **(b)** a [2]catenane comprising two interlocked macrocycles, necessitating breaking of a chemical bond in order to separate the two rings.

Nomenclature system for naming rotaxanes was first proposed[254] by Schill in 1971. In essence, the name of an interlocked species begins with a number in square brackets, indicating the total number of components comprising each individual architecture. Therefore, the catenane and rotaxane examples presented in **Figure 6.1** are both assembled from two components, and are thus named [2]catenane and [2]rotaxane. More recently, a new strategy to naming mechanically-interlocked architectures has been put forward[255,256] by Vögtle and co-workers and later also by IUPAC.

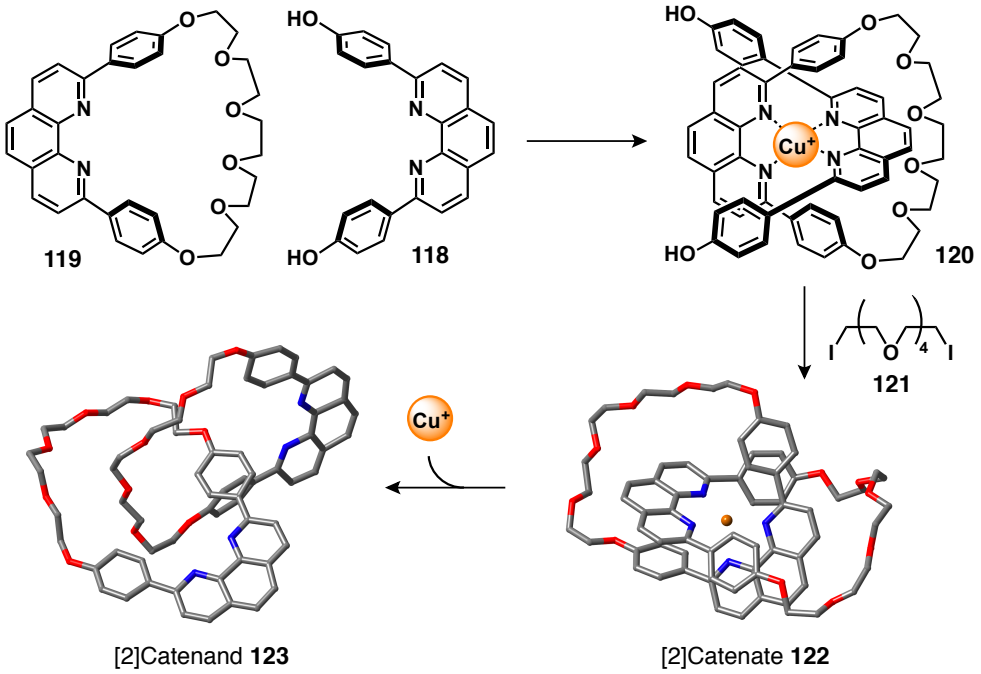
Initial attempts to synthesise MIAs were based on statistical approaches, that is synthetic procedures where these architectures were formed by chance during the cyclisation of a macrocyclic component. The first synthesis and isolation of a [2]catenane based on long-chain cycloalkanes (C_{34}) was reported[257] in 1960 by Wasserman, in 1% yield. Few years later, Harrison and Harrison reported[258] the first statistical synthesis of a [2]rotaxane (referred to as hooplane). The authors employed a clever, albeit extremely labour-intensive means of synthesising the desired architecture—the macrocyclic component **113** was attached to a solid support resin (**Scheme 6.1**), while the thread **114** and triphenylmethyl chloride **115** stoppering element were introduced in the the eluent. By repeated cycles of passing the solvent through the resin (up to 70 times), followed by hydrolysis of the support-bound architecture **116**, lead to the isolation of [2]rotaxane **117** in 6% yield.



Scheme 6.1 First [2]rotaxane synthesis as described by Harrison and Harrison in 1967. A macrocycle attached to a solid support **113** is repeatedly exposed to eluent containing thread **114** and triphenylmethyl chloride **115** as the stoppering element, resulting in support-bound rotaxane **116**. Hydrolysis of the bound structure produced [2]rotaxane **117** in 6% yield.

Clearly, statistical approaches to synthesis of MIAs permit isolation of these architectures, but in very poor yields. Over time, methods exploiting directed synthesis,[254,259] *i.e.* methods where the components to be mechanically-interlocked are covalently bound during the synthesis, began to appear. The field of mechanically-interlocked architectures, however, was revolutionised in 1983 following a report by Sauvage and co-workers introducing[260] a new class of metallo-catenane compounds. In this seminal

work, the authors exploited the interaction between a 2,9-diphenyl-1,10-phenanthroline ligand **118** and a copper cation (**Scheme 6.2**). Specifically, Cu^+ acted as a template for the assembly of a single phenanthroline ligand **118** and a macrocycle **119** incorporating this ligand. Reaction of this complex **120**, comprised of a macrocycle **119** threaded over the ligand **118** (also known as a pseudorotaxane) with spacer **121** produced [2]catenate[a] **122**—an interlocked architecture still associated with the Cu^+ ion used to template its synthesis, as evidenced[261] by its structure in the solid state. Upon removal of the metal template, the authors were able to isolate and analyse by X-ray diffraction the desired [2]catenand **123** (**Scheme 6.2**).



Scheme 6.2 Metal-templated synthesis of a catenand **123** reported by Sauvage and co-workers. A phenanthroline ligand **118** and a macrocycle **119** assemble around Cu^+ to produce a pseudorotaxane **120**. Reaction of complex **120** with a glycol spacer **121** produces a [2]catenate **122**, still coordinated to the Cu^+ ion, as evidenced[261] by its structure in the solid state. Demetallation produces the desired metal-free [2]catenand **123**, now with a dramatically altered molecular geometry. Scheme adapted from Ref. 261.

The term pseudorotaxane, as employed by Sauvage and co-workers in the synthesis of their catenand, denotes a general type of architecture where a macrocycle is threaded over a linear component lacking any stoppering groups. This precursor is suitable not only for the synthesis of catenanes but also rotaxanes. In both cases, however, the proportion of the pseudorotaxane relative to the uncomplexed species affects directly

[a]Mechanical-interlocking of two rings generates a [2]catenane. However, if a metal template is employed in the synthesis of the catenane, the resulting structure, still incorporating the metal ion is referred to as [2]catenate. Removal of the metal ion used as template produces an architecture termed as [2]catenand. Whilst the final structure is identical to the corresponding [2]catenane, the term is used to reflect the process of metal templation employed in its synthesis.

the quantity of a mechanically-interlocked architecture produced. Several years after the first report of a metallo-catenane, Gibson and co-workers have exploited[262] the same approach for the preparation of a [2]rotaxane. In fact, the authors utilised the same pseudorotaxane **120**, the reaction of which with bulky stoppering groups produced the target [2]rotaxane **124** in 42% yield (**Figure 6.3**)—a significant improvement relative to statistical methods.

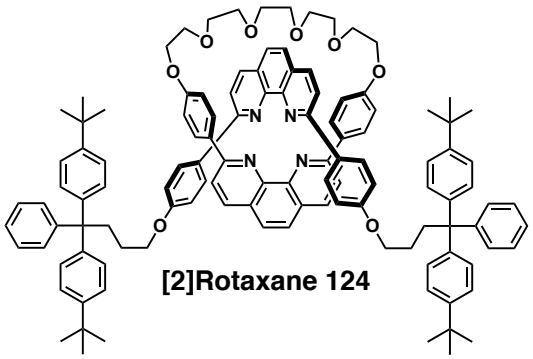


Figure 6.3 [2]Rotaxane **124** prepared by Gibson and co-workers using metal-templated synthesis and phenanthroline-based ligands.

Since its conception, metal template-mediated synthesis has been exploited[263,264] for the preparation of a variety of molecular architectures. These templating methods continued to evolve,[265] allowing synthesis of mechanically-interlocked architectures with templating methods based on the non-covalent molecular recognition toolkit (hydrogen bonding, halogen bonding, π–π stacking and hydrophobic interactions). These methods, illustrated using an example synthesis of a [2]rotaxane, can be grouped into three main categories: **(a)** stoppering method, **(b)** clipping approach and **(c)** slippage (**Figure 6.4**). The stoppering method takes place in two steps—the macrocycle first associates with the linear component *via* complementary recognition pattern to form a pseudorotaxane complex. The reaction of this pseudorotaxane with two equivalents of the stoppering component locks the macrocycle on the linear component, producing the target [2]rotaxane. In the clipping approach, the linear component is already equipped with bulky stoppers, and, instead, the macrocycle is the component that needs to be formed. Molecular recognition preorganises the smaller components required for the formation of the ring around the binding site on the linear thread, facilitating their reaction to form the desired [2]rotaxane. The slippage approach requires no bond forming steps, relying on size complementarity between the macrocyclic component and the stoppers. Specifically, given sufficient thermal energy, a macrocycle of a suitable size slips over the stoppers on to the linear component, where an appropriately designed binding site provides stabilising interactions in the formed [2]rotaxane. Whilst this

method does not template [2]rotaxane formation in the traditional sense[b], the binding interaction between complementary recognition sites on the macrocycle and the linear component together with the constrictive binding are crucial for generation of a stable interlocked rotaxane.

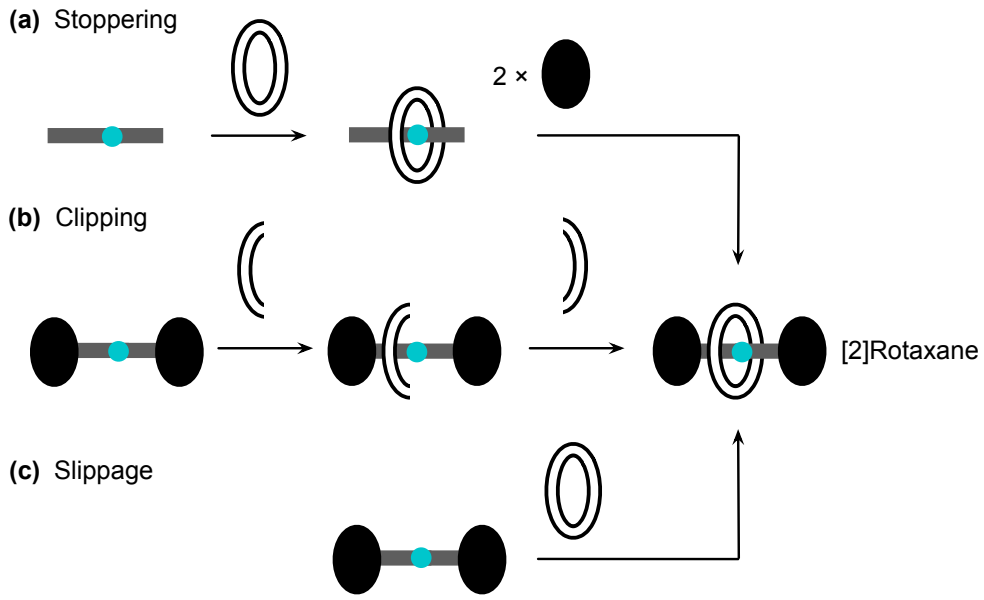
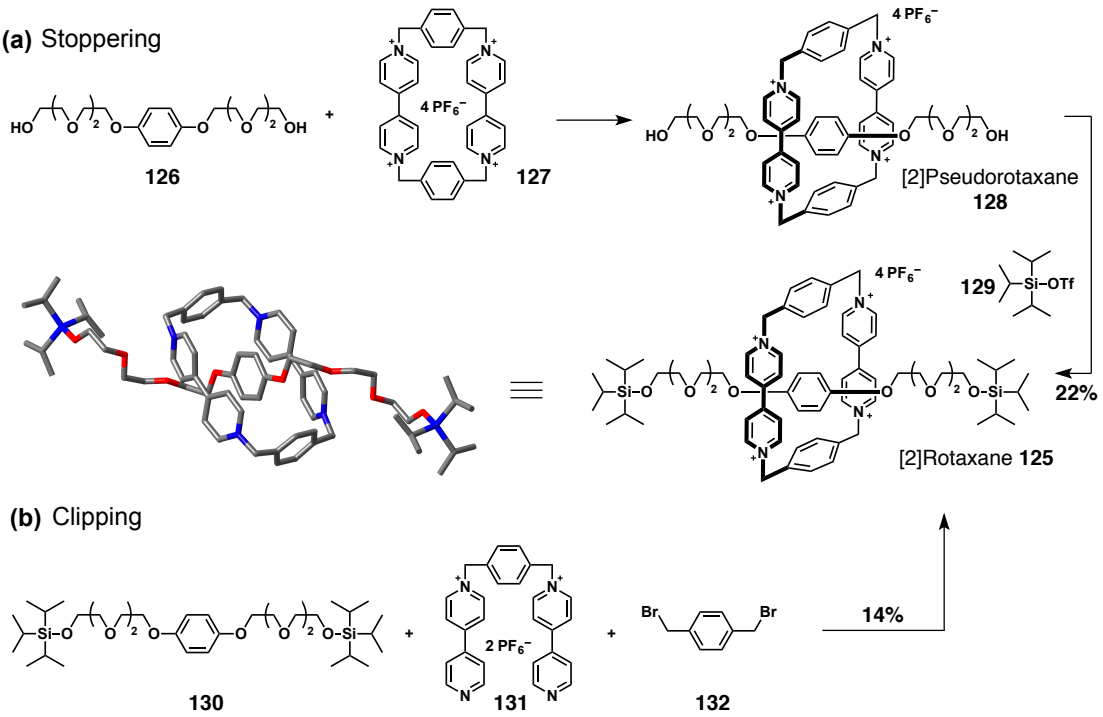


Figure 6.4 Three template-directed syntheses of [2]rotaxane, driven by non-covalent molecular recognition: **(a)** stoppering, **(b)** clipping and **(c)** slippage.

Stoddart and co-workers have demonstrated[271] formation of a [2]rotaxane using template based strategies exploiting noncovalent π–π stacking and charge transfer interactions. The authors achieved the synthesis of [2]rotaxane **125** using both the stoppering (**Scheme 6.3a**) and the clipping method (**Scheme 6.3b**). To form the rotaxane *via* the stoppering approach, linear component **126** was mixed with cyclobis(paraquat-*para*-phenylene) **127** to form [2]pseudorotaxane **128**. The reaction of this complex with two equivalents of stoppering component **129** produced the desired [2]rotaxane **125** in 22% yield. The capping method utilised threading component **130** already bearing bulky

[b]In a review published in 2010, Leigh and Hänni argue[266] that structures formed *via* the slippage mechanism are not formally rotaxanes but instead pseudorotaxanes, that exist as kinetically-stable structures only under a particular range of conditions. The authors reason that because these structures contain end groups that have clearly permitted the macrocycle to slip over them during the synthetic process, they do not qualify as rotaxanes, which are defined[267] as "molecules in which a ring encloses another, rod-like molecule having end-groups too large to pass through the ring opening and thus holds the rod-like molecule in position without covalent bonding". Making the distinction[268–270] between a pseudorotaxane and a rotaxane might not always be a straightforward task. Nevertheless, the argument put forward by Leigh and Hänni implies that the on and off mechanism for the slippage is identical. Another possible counterargument is that even rotaxanes formed using other synthetic approaches exhibit kinetic stability under certain conditions, *i.e.* if a rotaxane formed through a capping mechanism is given sufficient thermal energy, the rotaxane structure might likewise result in the dethreading of the macrocycle from the linear component.

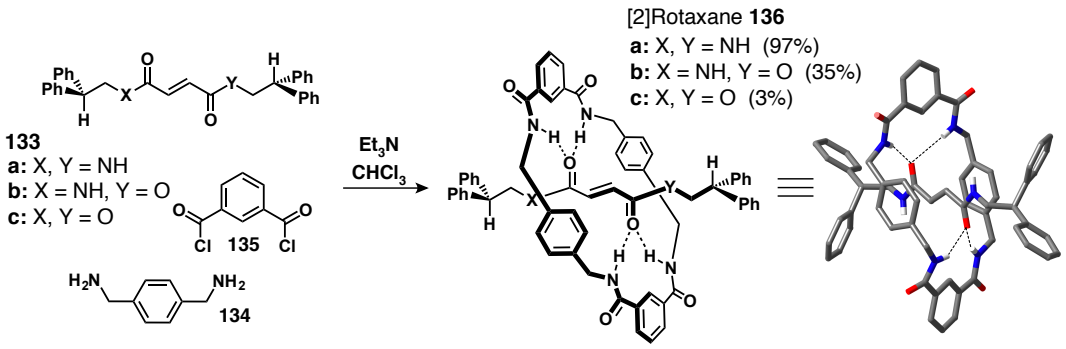
end groups. The reaction of thread **130** with bipyridinium dication **131** and stopper **132** produced [2]rotaxane **125** again, albeit in a slightly lower 14% yield.



Scheme 6.3 [2]Rotaxane **125** synthesised by Stoddart and co-workers using the **(a)** stoppering approach and **(b)** clipping method, exploiting non-covalent π–π stacking and charge transfer interactions. **(a)** The linear component **126** associates with cyclobis(paraquat-*para*-phenylene) **127** to form [2]pseudorotaxane **128**, which reacts with two equivalents of the stoppering component **129**. [2]Rotaxane **125** produced *via* this two-step stoppering approach was isolated in 22% yield. **(b)** In the capping approach, the threading component **130** already bearing the bulky stoppers was reacted with bipyridinium dication **131** and 4-bis(bromomethyl)benzene **132**, producing the [2]rotaxane **125** in 14% yield. Hydrogen atoms and counterions were omitted from the X-ray crystal structure of [2]rotaxane **125** for clarity. Scheme adapted from Ref. 271.

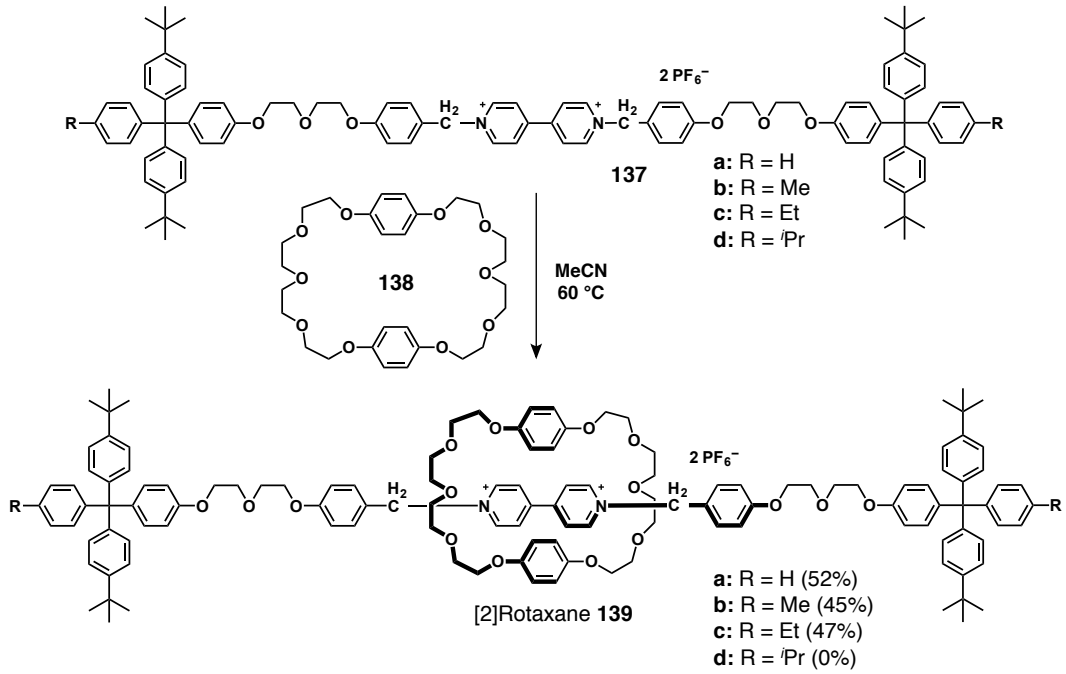
In a more recent example of a [2]rotaxane synthesis employing the clipping approach, Leigh and co-workers have demonstrated[272] the beneficial effect of preorganisation and structural rigidity on the efficiency of rotaxane synthesis. The authors investigated the reaction between three variations of thread **133** (**a** to **c**), containing hydrogen bonding motifs of various strengths (**Scheme 6.4**) and four equivalents of xylylene diamine **134** and isophthaloyl dichloride **135** to give the corresponding [2]rotaxanes **136a** to **136c**. The five component clipping reaction employing the fumaramide thread **133a** proceeded with a remarkable 97% yield—a value markedly higher than yields achieved in similar fumaramide-based rotaxane systems investigated[273,274] by Vögtle and co-workers. Following analysis of the structure of [2]rotaxanes **136a**–**c** (**Scheme 6.4**) in the solid state, the authors were able to attribute the effectiveness of the fumaramide thread **133a** as a template relative to the other threads (**133b** and **133c**) employed to

the arrangement of the two cooperative, amide-binding sites in a manner promoting a low-energy chair conformation of the macrocycle and its precursor around the thread. Compared to a simple dipeptide motif that can likewise be utilised as a template for the formation of a benzylic macrocycle, this thread is significantly more rigid, and thus, no internal degrees of freedom are lost upon complexation.



Scheme 6.4 Hydrogen-bonding mediated recognition directs the formation of a [2]rotaxane from a five-component reaction mixture *via* the clipping method. Reaction of three different thread **133** frameworks (**a** to **c**) with four equivalents of xylylene diamine **134** and isophthaloyl dichloride **135**. Of the three [2]rotaxanes**136a-c** examined, the architecture based on the fumaramide thread **133a** was formed most efficiently (97%). Crystal structure of the [2]rotaxane **136a** revealed the favourable low-energy chair conformation of the formed macrocycle and the hydrogen-bonding between the thread and the ring (dashed black lines). Scheme adapted from Ref. 272.

The slippage mechanism was first described[275] by Stoddart and co-workers in 1993, (**Scheme 6.5**). The authors probed the slippage pathway by employing thread **137**, containing a 4,4'-bipyridinium dication equipped with increasingly bulky stoppers (**a** to **d**). Heating each thread in the presence of four equivalents of bis-*para*-phenylene-34-crown-10 macrocycle **138** for ten days revealed formation of [2]rotaxanes **139a-c** in 52%, 45% and 47% yield, respectively. No rotaxane formation was detected in the experiment employing thread **137d** with the bulkiest R group (^{i}Pr), even though this system was likewise capable of forming the stabilising interactions between the macrocycle phenylene rings and the bipyridinium motif on the thread. Subsequent computational analyses undertaken by Stoddart and co-workers on this system showed[276] a good agreement with the experimentally-derived data—revealing that the formation of [2]rotaxane **137d**, bearing the bulky ^{i}Pr stoppers, is associated with a significantly higher energy barrier than any of the other threads. In this work, the authors demonstrated successfully for the first time, that a mechanically-interlocked architecture can be produced at elevated temperatures by exploiting size complementarity between the stoppering groups and the macrocyclic ring, followed by the thermodynamic trapping of the macrocycle in the formed architecture *via* non-covalent interactions.



Scheme 6.5 Heating of thread **137a**-**137d** in the presence of macrocycle **138** in MeCN for 10 days allowed formation of [2]rotaxane **139** (**a** to **c**) *via* the slippage method in 52%, 45% and 47% yield, respectively. Thread **137d** bearing the ^{i}Pr group on the stoppers proved to be too large for the macrocycle to slip over at this temperature. Scheme adapted from Ref. 275.

More recently, metal templating strategies have been elaborated[266,277–281] to what has been termed as the active-template approach, where the metal used in the synthesis of the interlocked architecture not only preorganises the reaction components but also takes part in the reaction itself as a catalyst. The first example employing this method was presented[282] by Leigh and co-workers. The authors designed a macrocycle **140** (**Figure 6.5**, black) capable of coordinating Cu^{+}. The metal catalyses the reaction between alkyne **141** (**Figure 6.5**, blue) and azide **142** (**Figure 6.5**, red), whilst simultaneously ensuring the correct position of these fragments so as to produce the target [2]rotaxane **143**. Following optimisation of the reaction conditions, the authors were able to synthesise [2]rotaxane **143** in 82% yield, using only catalytic amount of the catalyst (4 mol%)—dramatically reducing the quantity of metal template necessary for synthesis of a MIA.

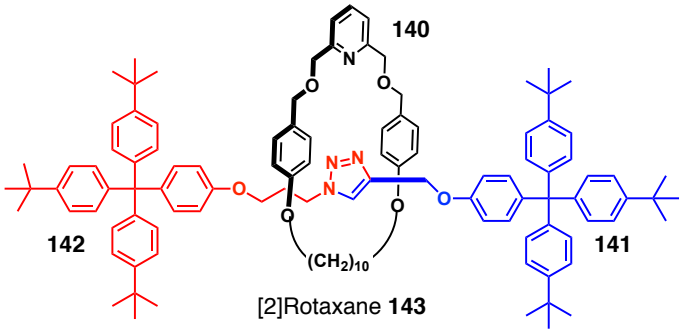


Figure 6.5 [2]Rotaxane **143** produced[282] by Leigh and co-workers using active metal template strategy. Copper catalyses the click reaction between alkyne **141** (blue) and azide **142** (red) in the presence of macrocycle **140** (black) incorporating a pyridine unit that directs the reaction through its cavity.

6.2.2 Rotaxanes: applications

Rotaxanes contain two or more mechanically-interlocked components, held together not through a covalent link but instead through a mechanical bond. Movement of the macrocyclic component on the thread can be limited by the presence of the stoppering groups and the stabilising non-covalent interactions between the macrocycle and thread. Nevertheless, it is possible for the macrocycle to move[283] relative to the threading component *via* two types of motions: pirouetting and translational motion. Pirouetting describes the random Brownian rotational movement of the macrocycle on the ring (**Figure 6.6**, blue), whereas translation represents the movement of the ring along the thread itself (**Figure 6.6**, red). As a result of these motions, the components making up a rotaxane architecture can adopt various co-conformations.

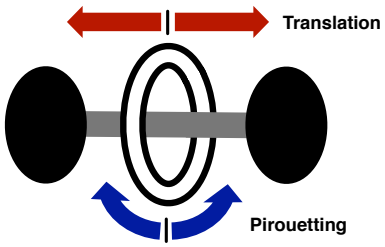


Figure 6.6 Cartoon illustrating [2]rotaxane and the corresponding translation (red arrows) and pirouetting (blue arrows) motions of the macrocycle ring component relative to the thread within this architecture.

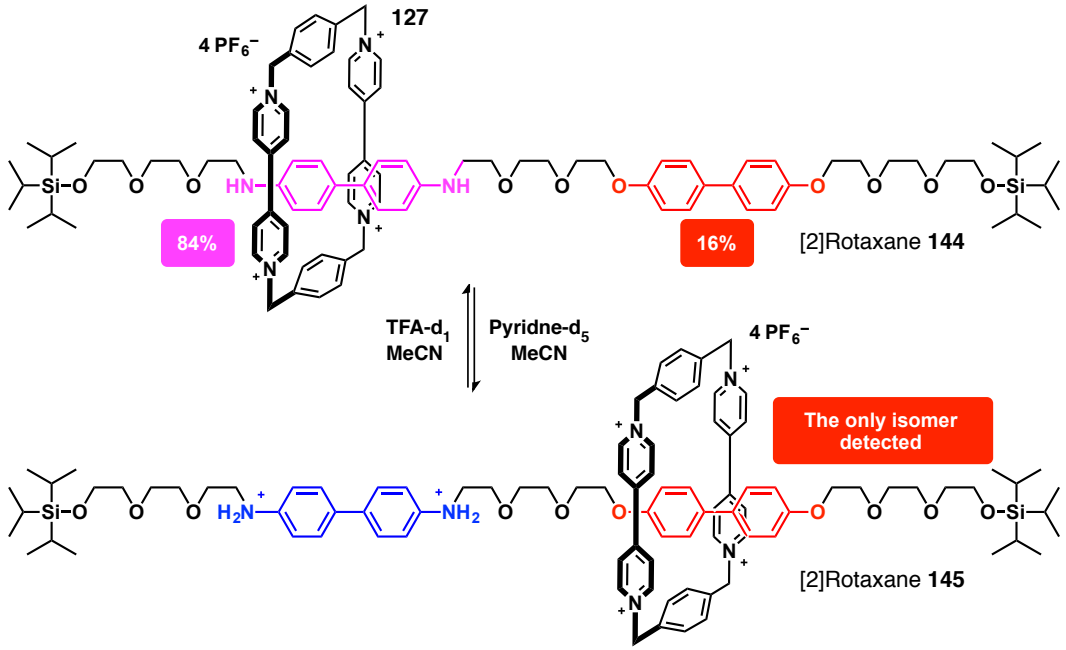
It could be envisaged that if the threading component possessed two binding sites, the so-called "stations",[283] the macrocycle could shuttle between them. In a situation where the two binding sites are identical, *i.e.* degenerate, the two stations will be populated equally by the macrocycle. Any motion of one component relative to another, be it through pirouetting, shuttling, *etc.* necessitates transient breaking of the interactions between the macrocycle and its binding site on the thread—a process associated with a

reaction barrier, dependent on the strength of the interactions taking place. For example, for the three fumaric-acid-based rotaxanes **136a** to **c** (**Scheme 6.4**), the reaction yields correlated[272] inversely with the rate of pirouetting—the stronger the hydrogen bond binding motif on thread, the lower the rate of rotation around the thread. Similar dependency was found to hold for the rate of shuffling. as well

The first example of a molecular shuttle was reported[284] by Stoddart and co-workers in 1991. The rotaxane shuttle was formed on a symmetrical thread, very similar to thread **130** described[271] in **Scheme 6.3**, but containing an additional hydroquinone binding site and the previously employed macrocycle **127**. With this prototype example of a degenerate molecular shuttle, the authors opened[284] up the possibilities to "*desymmetrise the molecular shuttle by inserting non-identical 'stations' along the polyether 'thread' in such a manner that these different 'stations' can be addressed selectively by chemical, electrochemical, or photochemical means and so provide a mechanism to drive the 'bead' to and fro between stations along the 'thread'.*"

Stoddart and co-workers have extended their work to produce[285] a molecular shuttle capable of switching between two different stations in response to changes in pH and electrochemical input. The threading component was designed to incorporate a benzidine (**Scheme 6.6**, purple) and bisphenol (**Scheme 6.6**, red) recognition units, capable of interacting with macrocycle **127** (**Scheme 6.6**). In the absence of any input, *i.e.* at equilibrium (in MeCN), the benzidine station was found to be more populated (84%) than the biphenol site (16%) in [2]rotaxane **144**. Addition of TFA-d_1 to the system acts as a chemical input, driving the switch from the now protonated benzidine station (**Scheme 6.6**, blue) to the bisphenol motif in [2]rotaxane **145**. By showing that the switching process is reversible and responsive to chemical input, the authors demonstrated the first example of what has become known as a molecular machine.[286,287]

The interest in switchable, mechanically-interlocked architectures continued to increase over time, with numerous new examples appearing[288–293] in the literature. The next advance in the field involved the development of systems where the motion of one component relative to another in an interlocked structure produces a change that can be harnessed to express function or work. One of the avenues that garnered significant attention is the development of molecular machines for the use in molecular electronics,[294,295] led mainly by the laboratories of Stoddart and Heath.



Scheme 6.6 A stimuli-responsive [2]rotaxane molecular shuttle developed by Stoddart and co-workers. [2]Rotaxane **144** is comprised of macrocycle **127** and a threading component equipped with a benzidine (purple) and a bisphenol (red) station. At equilibrium in MeCN, the benzidine station is the preferred binding site for the macrocycle (84%). Upon addition of TFA-d_1, the benzidine station is protonated (purple), producing [2]rotaxane **145** which binds the macrocycle preferentially at the bisphenol station. Scheme adapted from Ref. 285.

In 2007, their work on molecular switches[285,296,297] culminated in the development[298] of a 160000-bit molecular electronic memory circuit, incorporating a bistable [2]rotaxane switch **146** (**Figure 6.7**). The threading component of the rotaxane was equipped with an electron rich tetrathiafulvalene (TTF) site and a 1,5-dioxynaphthalene (DNP) site capable of associating with an electron poor macrocycle **127**. The circuit was constructed from a layer of [2]rotaxanes **146**, sandwiched between Si and Ti electrodes, with a hydrophilic stopper (light blue) oriented towards the Si electrode. Macrocycle bound at the TTF station equals to the low conductance State 0 (**Figure 6.7**, ground co-conformation). Oxidation of the TTF station (**Figure 6.7**, green) forces the macrocycle **127** to the red DNP. The reduction of the TTF radical cation to the neutral TTF produces a high conductance metastable State 1 (**Figure 6.7**, metastable co-conformation), which relaxes back to the ground state with half life of around 60 minutes within the device.

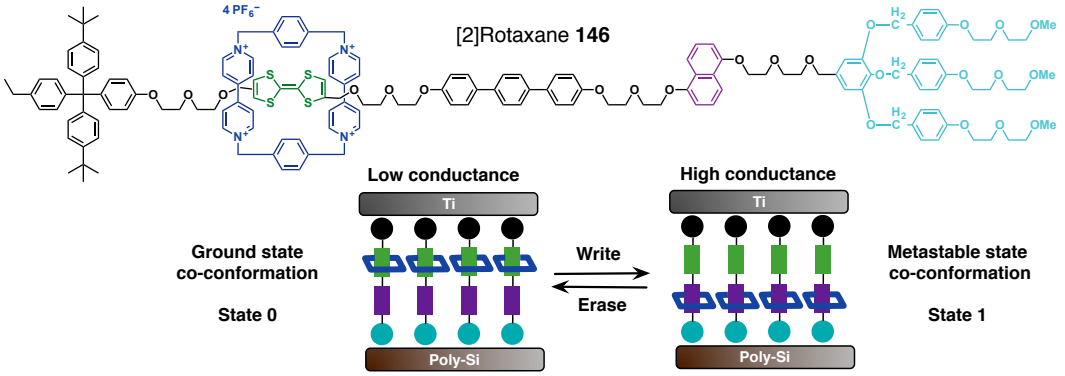


Figure 6.7 A bistable [2]rotaxane **146** was incorporated by Stoddart, Heat and co-workers into a functional 160000-bit crossbar memory device. The rotaxane contains an electron rich tetrathiafulvalene (TTF, green site) and a 1,5-dioxynaphthalene (DNP, red site) stations for macrocycle **127**. A layer of [2]rotaxane **146**, sandwiched between Si and Ti electrodes goes through cycles of low and high conductance mediated by the oxidation and reduction of the TTF station. Figure adapted from Ref. 244.

In 2013, Leigh and co-workers have reported[299] an interesting example of a molecular machine capable of performing work, in this case, sequence specific peptide synthesis. This molecular machine, based on [2]rotaxane **147** (**Figure 6.8**) incorporating a macrocycle **148** equipped with a reactive arm capable of reacting with the amino acid building blocks distributed along the thread of the rotaxane. Following each native chemical ligation reaction, the catalytic thiol site was restored and the altered macrocycle could move further along the thread. Following the reaction with the last amino acid residue (**Figure 6.8**, purple), the macrocycle, now bearing three extra amino acid residues can slip of the thread. Finally, hydrolysis released the completed hydrazide peptide **149** from the macrocycle.

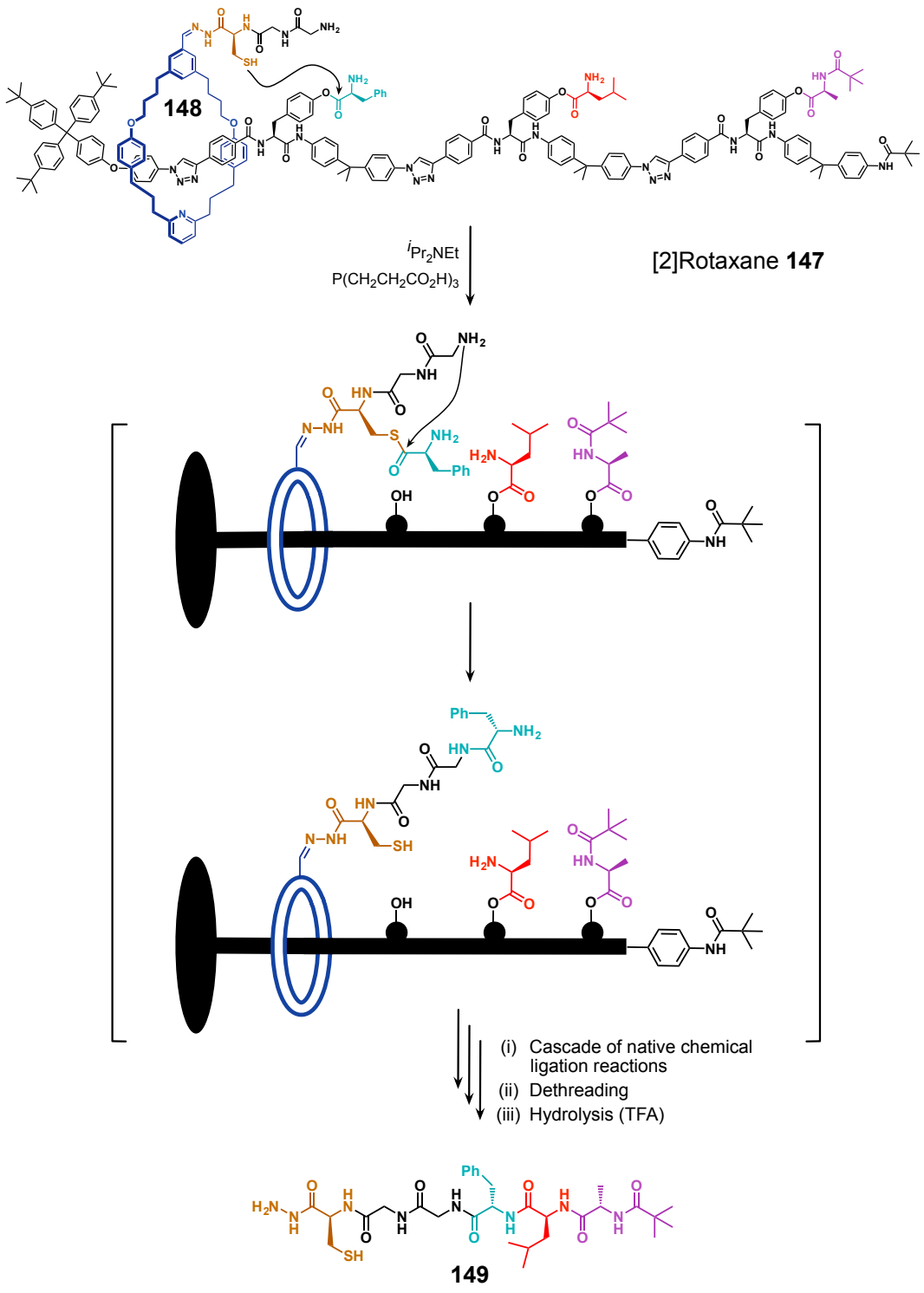


Figure 6.8 [2]Rotaxane **147** molecular machine designed[b]y Leigh and co-workers. The macrocycle **148** bearing a reactive arm travels along the threading component taking part in a sequence of native chemical ligation reactions. After the macrocycle reacts with the last amino acid (purple), dethreading and subsequent hydrolysis releases the hydrazide peptide **149** from the macrocycle. Figure adapted from Ref. 299.

6.3 Integrating [2]rotaxane assembly with replication processes

This short introduction to the field of mechanically-interlocked architectures illustrates the considerable progress that has been achieved[292,293,295,300] over the last 50 years. Synthetic techniques, exploiting the power of templates and molecular recognition for the formation of mechanically-interlocked architectures have become significantly more efficient, providing access to a variety of intricate and functional architectures. In the next section, rotaxane formation will be examined in a system where the recognition-mediated processes required for its formation operate simultaneously with template-directed replication processes.

6.3.1 Structural requirements and kinetic models

The minimal model of self-replication was introduced in **Chapter 1**—in order to self-replicate, a molecule must be equipped with complementary reactive sites (**Figure 6.9a**) and recognition sites (**Figure 6.9b**). In order to be able to template its own synthesis and manage simultaneously the construction of a [2]rotaxane, additional recognition sites are required to mediate the association of a macrocycle with the threading component (**Figure 6.9c**). Association of the macrocycle on the thread produces a [2]pseudorotaxane, a complex essential for the successful formation of a self-replicating [2]rotaxane.

There are three obvious ways to organise the recognition and reactive sites in a system incorporating replication processes with rotaxane assembly. These three models (Model 1 to 3) for a self-replicating rotaxane system, exploiting the two-stage stoppering method, can be identified based on the relative orientation and arrangement of the recognition and reactive sites (**Figure 6.9c**), designed to probe the interplay between the two recognition-instructed processes, namely, replication and [2]rotaxane formation.

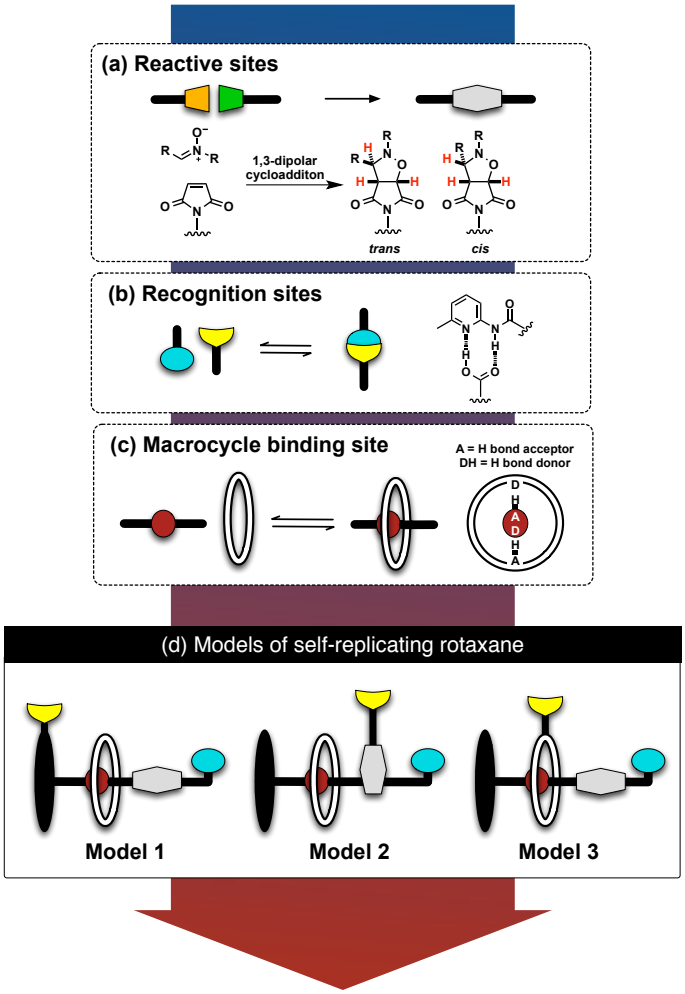


Figure 6.9 **(a)** Reactive and **(b)** to **(c)** recognition elements required for the formation of a self-replicating rotaxane. **(d)** Three models for the integration of self-replication with the formation of a rotaxane. Model 1 incorporates the macrocycle binding site within the self-replicating core. Model 2 possess the recognition site required for macrocycle binding away from the reactive and recognition components needed for replication. In Model 3, one of the recognition site utilised in self-replication is located on the macrocycle itself.

The main feature of Model 1 is the position of the macrocycle binding site (red), which is located within the self-replicating core of the molecule. The model is governed by a central equilibrium (**Figure 6.10**, centre) between a linear component **L**, which possesses the binding site required for association with a macrocycle, and macrocycle **M** itself. Both the pseudorotaxane complex [**L**·**M**] and the linear component possess also a reactive site, allowing the reaction with a stoppering component **S**. These reactions produce two products, a thread **T**, lacking the macrocycle, and a [2]rotaxane **R**. Once formed, both **T** and **R** can assemble with the unreacted reaction components into catalytically-active ternary ([**T**·**L**·**S**]) and quaternary ([**R**·**L**·**M**·**S**]) complexes, which enable them to template their own formation (**Figure 6.10**, left), in a manner similar to that described for the minimal model of self-replication in **Chapter 1**.

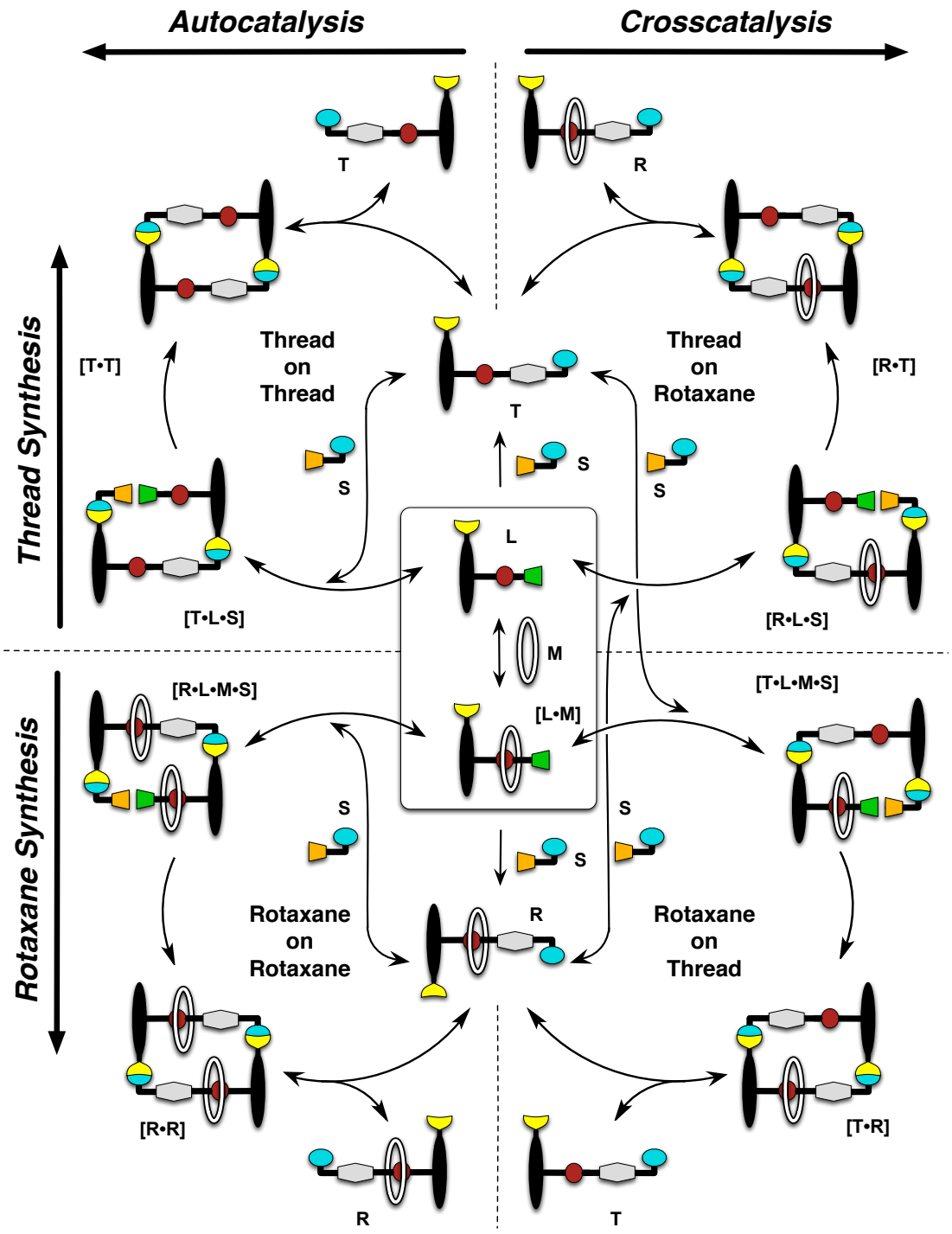


Figure 6.10 Cartoon representation of Model 1. The replicating system is based around a key equilibrium (solid rectangle, centre) between a linear component **L** and macrocycle **M**, to produce a pseudorotaxane complex [**L**·**M**]. Following the reaction of **L** and [**L**·**M**] with stopper **S**, the network diverges from the central equilibrium into separate, yet interconnected, auto- (left) and crosscatalytic (right) thread **T** (top) and [2]rotaxane **R** (bottom) forming cycles. For simplicity, equilibrium arrows are shown as double-headed arrows. The yellow and blue cartoons represent the recognition sites and green and blue the reactive sites needed for self-replication. Grey shape represents the reaction product and the macrocycle-binding site is shown in red.

As a result of the identical nature of the recognition sites present in both the thread and the rotaxane, these templates can theoretically also take part in crosscatalytic cycles, driven *via* the formation of [**L**·**S**·**R**] and [**L**·**M**·**S**·**T**] complexes (**Figure 6.10**, right). However, the location of the macrocycle on the linear component within this model, *i.e*. directly between the recognition sites required for self-replication, means that the interplay between the macrocycle and the additional reaction and recognition processes could diminish the ability of the thread and rotaxane templates to perform in the the crosscatalytic pathways.

In contrast to Model 1, Model 2 (**Figure 6.11**) represents a system where the recognition site mediating macrocycle binding is positioned at the periphery of the structure, away from the recognition and reactive elements required for self-replication. Therefore, the position of the macrocyclic component within the structure in Model 2 is more likely to leave the replication processes unaffected than the macrocycle location in Model 1. Specifically, the position of the macrocycle should ensure that the reactivity of the pseudorotaxane complex [**L**·**M**] is very similar, if not identical to that of the linear component itself. Similarly, the distance of the macrocycle binding site from the recognition sites, driving replication processes in the template structure should permit the thread and the rotaxane to take part in auto- and crosscatalytic cycles with comparable efficiency.

Model 3 (**Figure 6.12**) provides a markedly different scenario, where one of the recognition sites required for self-replication is located on the macrocycle itself. The orientation of the recognition and reactive elements means that the macrocycle is directly involved in the replication—making the outcome of this interplay between the recognition and reaction processes on the efficiency of rotaxane formation extremely challenging to predict. Nevertheless, Model 3 provides a scenario where the [2]rotaxane product is the only species in the system capable of utilising the template-driven processes to direct its own formation. The formation of the thread, lacking the macrocycle equipped with a recognition site can proceed only through the slow bimolecular pathway.

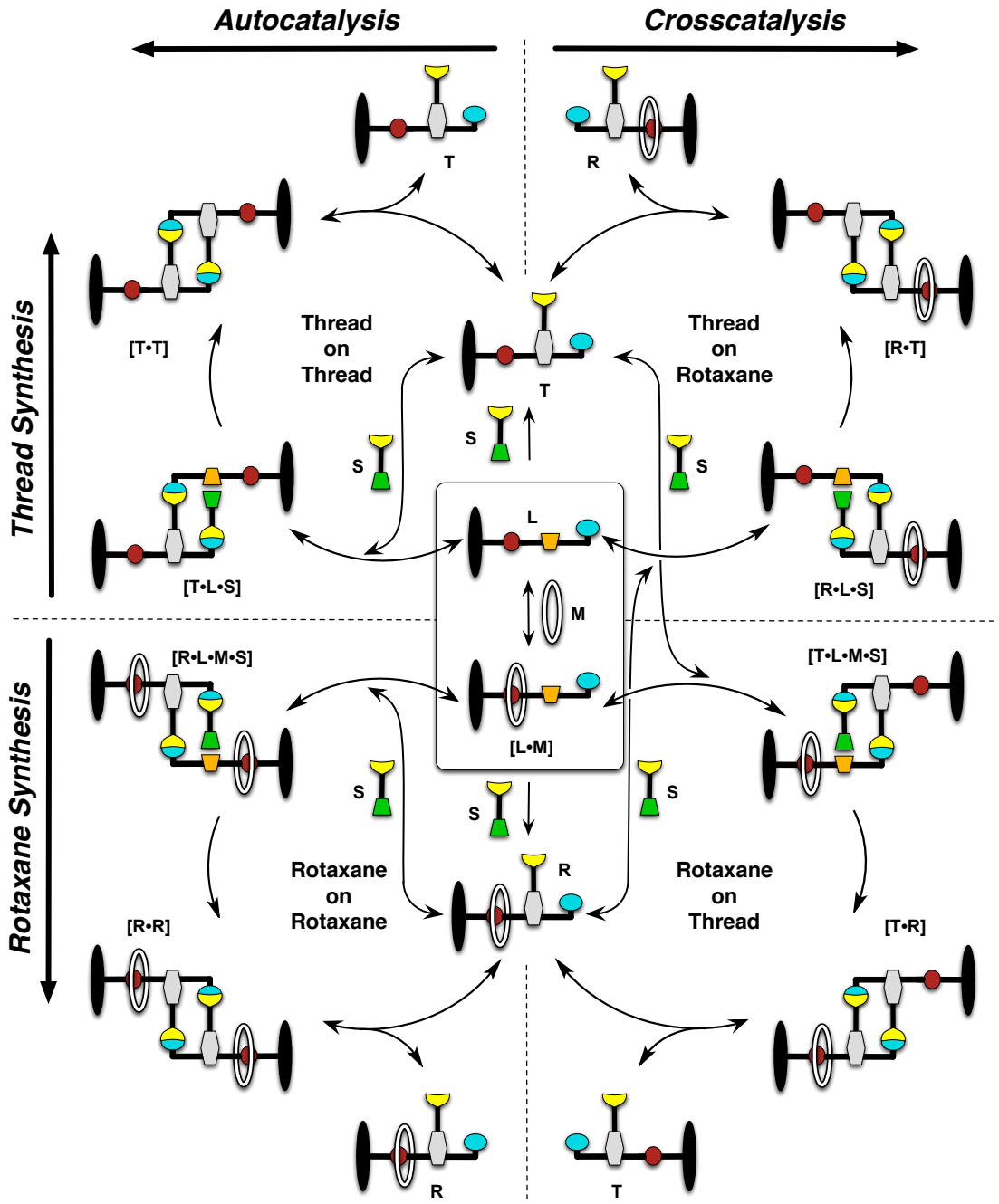


Figure 6.11 Cartoon representation of Model 2. A replicating network exploiting orthogonal recognition processes results in the formation of a [2]rotaxane **R** and thread **T**. The replicating system is based around a key equilibrium (solid rectangle, centre) between a linear component **L** and macrocycle **M**, to produce a pseudorotaxane complex [**L**·**M**]. From the central equilibrium, the network diverges into separate, yet interconnected, auto- (left) and crosscatalytic (right) thread (top) and rotaxane (bottom) forming cycles. For simplicity, equilibrium arrows are shown as double-headed arrows. The yellow and blue cartoons represent the recognition sites and green and blue the reactive sites needed for self-replication. Grey shape represents the reaction product and the macrocycle-binding site is shown in red. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

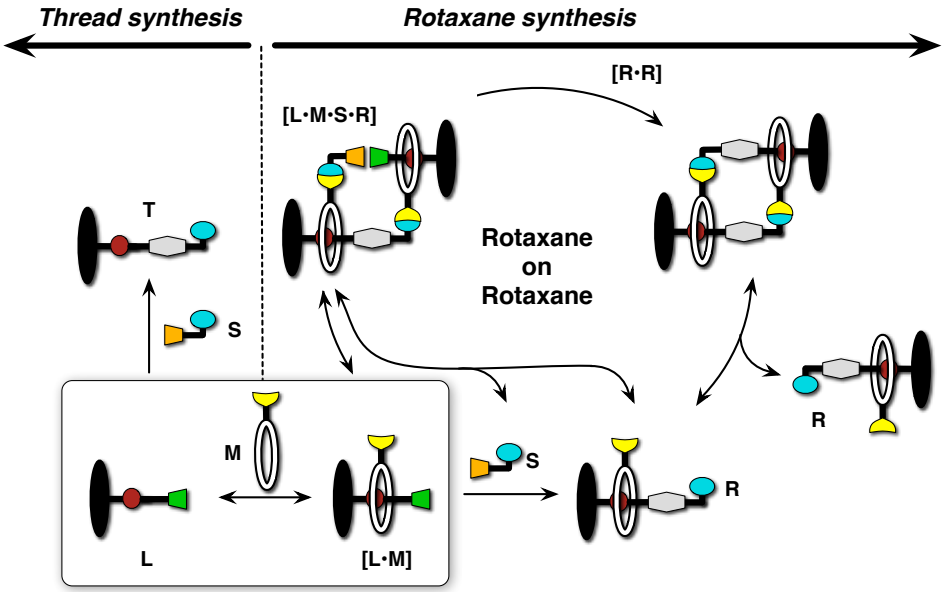


Figure 6.12 Cartoon representation of Model 3, where macrocycle **M** bears one of the recognition sites required for self-replication. The replicating system is based around a key equilibrium (solid rectangle, bottom left) between a linear component **L** and macrocycle **M**, to produce a pseudorotaxane complex [**L**·**M**]. From the central equilibrium, the linear component can react with stopper **S** only *via* template independent bimolecular pathway to produce thread **T**. The macrocycle bearing pseudorotaxane [**L**·**M**] can react with **S** through both the bimolecular reaction to form [2]rotaxane **R** and through recognition-mediated self-replicating pathway driven by catalytically-active complex [**L**·**M**·**S**·**R**]. No crosscatalytic pathways are possible in this network. For simplicity, equilibrium arrows are shown as double-headed arrows. The yellow and blue cartoons represent the recognition sites and green and blue the reactive sites needed for self-replication. Grey shape represents the reaction product and the macrocycle-binding site is shown in red.

6.3.2 Previous work: Model 1

The first attempt to integrate replication processes with the formation of a [2]rotaxane was described[302] by Vidonne and Philp in 2008. The network components (**Figure 6.13a**) were designed according to Model 1, with the macrocycle binding site located within the core of the replicating framework. Linear component **150** was designed to contain an amide binding site, in order to drive the hydrogen-bonding mediated association with a glycol spacer macrocycle (**GM**). The linear component was also equipped with a carboxylic acid site, in order to permit formation of the catalytically-active complexes, governed by the recognition of the complementary amidopyridine unit on the stoppering element **151** (**Figure 6.13a**). In the presence of macrocycle **GM**, both the linear component and the pseudorotaxane [**150**·**GM**] possessed the capacity to react with the stopper **151** *via* 1,3-dipolar cycloaddition reactions to produce *trans* and *cis* diastereoisomeric products. As a result of the position and orientation of the

recognition and reactive elements, only the *trans* diastereoisomer possesses the open conformation required for participation in template-directed reaction pathways.

Examination of the reaction between linear component **150** and stopper **151** in the absence of a macrocycle ($CDCl_3$, 25 °C, 20 mM) revealed that thread **152** had reached 49% conversion after eight hours, in a 7.9 ratio of [*trans*]/[*cis*] diastereoisomers. Analysis of the same reaction in the presence of one equivalent of macrocycle **GM**, where both thread and [2]rotaxane **153** can be formed, however, revealed a markedly different result. The combined conversion after eight hours dropped to 33% with a [rotaxane]/[thread] ratio of 0.45. In addition to lower overall conversion, the diastereoselectivity of thread formation decreased from 7.9 to 6 in favour of the *trans* product. Surprisingly, the [2]rotaxane **153** formed the *cis* diastereoisomer, incapable of templating its own formation preferentially ([*cis*]/[*trans*] ratio = 1.4).

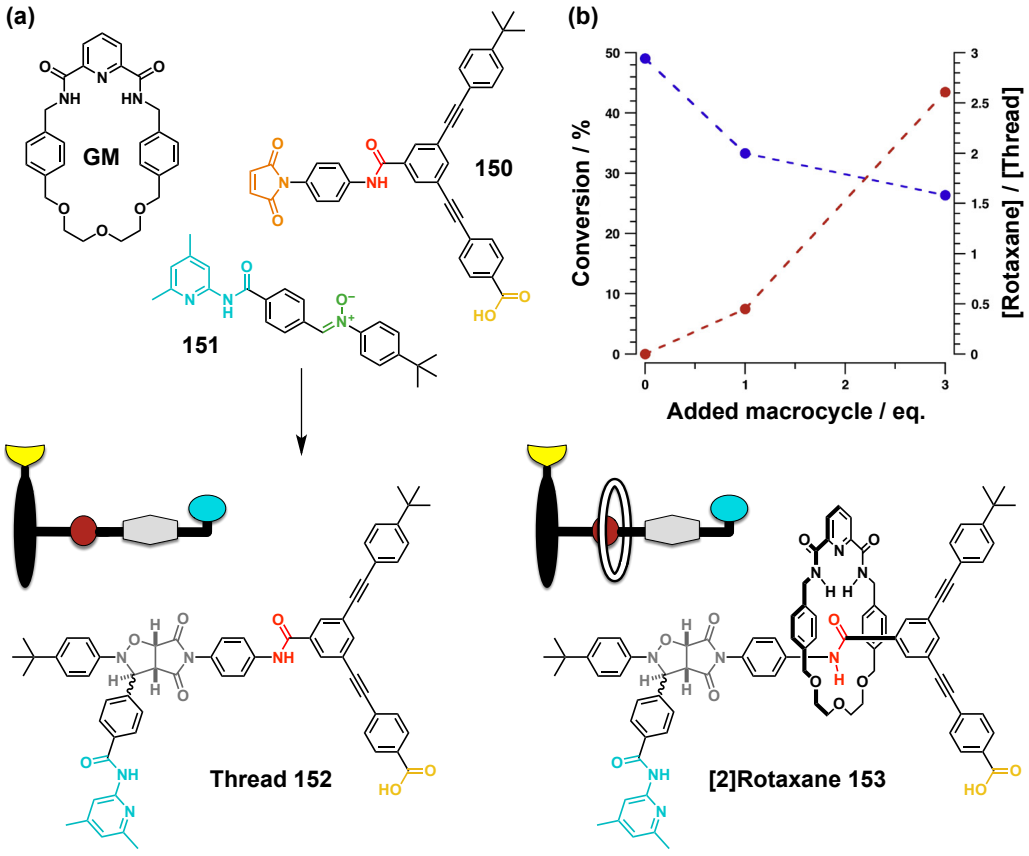


Figure 6.13 **(a)** Design of a replicating network integrating replication processes with the assembly of a [2]rotaxane *via* Model 1. Reaction of linear component **150** with stoppering element **151** in the presence of glycol macrocycle **GM** can produce two templates, a thread **152** lacking the macrocycle and a [2]rotaxane **153**. Green and orange represent the reactive sites, whilst yellow and blue denote the recognition sites required for self-replication. The macrocycle binding site is shown in red. **(b)** The effect of increasing the number of equivalents of the macrocycle **GM** in the reaction mixture on the conversion (blue circles) to cycloadduct products and the ratio (red circles) of [rotaxane]/[thread] formed in the system ($CDCl_3$ at 25 °C, 20 mM). Figure adapted from Ref. 302 with permission from Elsevier.

In order to circumvent the low reactivity of the pseudorotaxane and to achieve a higher proportion of the rotaxane in the system, the authors repeated the same kinetic experiment in the presence of three equivalents of macrocycle **GM** (**Figure 6.13b**). The experiments showed a significant increase in the ratio of [rotaxane]/[thread] to 2.6, however, with a drop in conversion to 26% after eight hours. The reaction processes leading to the formation of [2]rotaxane **153** continued to favour the formation of the *cis* diastereoisomer. While no template-instructed experiments were reported in this study, using a combination of kinetic simulation and computational modelling, the authors were able to attribute the lower reactivity of the linear component bound within the pseudorotaxane to a macrocycle-induced obstruction of the maleimide double bond. Similarly, the *cis* diastereoselectivity of the rotaxane was found to be affected by unfavourable interactions between a *tert*-butyl group on the stopper and the glycol chain of the macrocycles.

We have investigated[303] Model 1 further using an alternative design (**Figure 6.14a**). In this case, a significantly longer spacer was employed between the reactive maleimide site and the macrocycle binding site in the linear component **154** in order to avoid reduced pseudorotaxane reactivity as a result of the supramolecular steric effect upon macrocycle binding. The authors also employed a different macrocycle, **PM**, equipped with an additional pyridine unit, capable of stronger hydrogen-bonding mediated association with the amide in **154**. As in the previous design, the 1,3-dipolar cycloaddition reaction of linear component **154** with stopper **155** can produce thread **156** in the absence of macrocycle and both thread and [2]rotaxane **157** in its presence. In order to determine which catalytic pathways are active, the authors undertook a comprehensive kinetic analysis.

The kinetic analyses of thread in the absence of the macrocyclic component revealed that formation of thread **156** is extremely efficient and highly diastereoselective for the *trans* cycloadduct. Analysis of thread kinetic profile in the presence of preformed thread template confirmed the ability of the designed thread replicator to template its own formation, as evidenced by disappearance of the lag period. The reaction of linear component **154** with a stopper **155** in the presence of **PM** resulted in a [thread]/[roAVtaxane] ratio of 1.5. The formation of [2]rotaxane **157** was only slightly less efficient than that of thread and the reaction maintained fairly high diastereoselectivity—an improvement relative to the first design and higher selectivity for the *trans* diastereoisomer than that observed in non-recognition mediated, purely bimolecular, 1,3-dipolar cycloaddition reactions. Analysis of the reaction between the linear component, macrocycle and stopper in the presence of preformed template of either thread or rotaxane, however, showed no enhancement in the formation of [2]rotaxane **157**, indicating that the rotaxane was

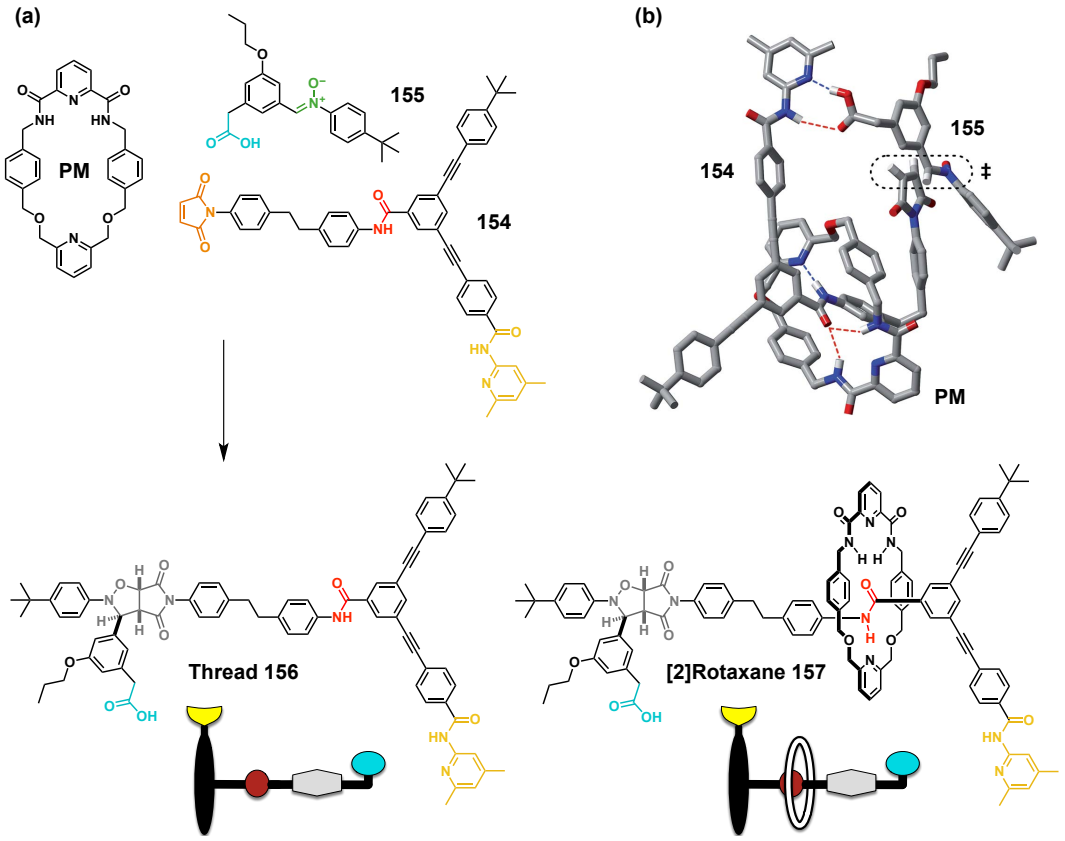


Figure 6.14 **(a)** Design of a replicating network integrating replication processes with the assembly of a [2]rotaxane *via* Model 1. Reaction of the linear component **154** with a stoppering element **155** in the presence of glycol macrocycle **PM** can produce two templates, a thread **156** lacking the macrocycle and a [2]rotaxane **157**. Green and orange represent the reactive sites, whilst yellow and blue denote the recognition sites required for self-replication. The macrocycle binding site is shown in red. **(b)** The effect of increasing the number of equivalents of the macrocycle **PM** in the reaction mixture on the conversion to cycloadduct products (blue circles) and the ratio of [rotaxane]/[thread] (red circles) formed in the system ($CDCl_3$ at 25 °C, 20 mM). Figure adapted from Ref. 303 with permission from the Royal Society of Chemistry.

not capable of templating its own formation nor was the thread a suitable crosscatalytic template. Thread formation only was shown to be catalysed by the presence of the preformed thread template, but not by the presence of the rotaxane template. A combination of kinetic fitting, kinetic simulations and computational modelling revealed that the presence of the flexible spacer, separating the maleimide reactive site and the amide recognition unit in the linear component **154**, brings about a degree of flexibility that permits the formation of the [2]rotaxane *via* a recognition-mediated, yet fairly inefficient ternary complex [**154**·**PM**·**155**] pathway (**Figure 6.14b**).

Investigations directed at embedding the assembly of a [2]rotaxane in a replicating network *via* the Model 1 clearly present a number of challenges. Binding of the macrocycle to the linear component adds an additional layer of complexity to a replicating system, with the potential to diminish the reactivity of the pseudorotaxane

complex [**L**·**M**] and thus also the capacity of a species to template its formation, as illustrated in the first example. Design of self-replicating systems presents a challenge—as demonstrated by the second example—structural flexibility can inadvertently promote formation of products *via* competing recognition-mediated but template independent pathways. Nevertheless, in both examples, the efficient formation of thread demonstrated, together with the lack of any crosscatalytic activity, that further optimisation of the design features can afford a network capable of forming a rotaxane architecture preferentially and efficiently through the template-catalysed autocatalytic pathway.

6.4 Model 2

The work presented in this chapter aims to explore the possibility of exploiting the orthogonal recognition processes of Model 2 in order to drive the replication and assembly of a [2]rotaxane. Elaborating on currently unpublished work[304] within the Philp laboratory, it is envisaged that Model 2 should permit formation of a replicating network where all catalytic cycles are active, as a result of the location of the macrocycle at the periphery of the molecular structure. An additional advantage of Model 2 is the possibility of employing a known highly efficient replicating framework as one of the stoppers, as a result of the orthogonal recognition processes.

6.4.1 Design and synthesis of rotaxane components

A rotaxane forming replicating system requires a linear component **L** and a stopper **S**, both equipped with complementary recognition and reactive sites. Reaction of these components should produce, in the absence of a macrocycle, a thread molecule **T** capable of efficient template-directed self-replication mediated by the assembly of the unreacted building blocks on the template in a ternary complex [**L**·**S**·**T**]. The reaction component **L** possesses an additional recognition site with complementary recognition motif to that of a macrocycle **M**. In the presence of a macrocycle, an equilibrium between the linear component in isolation and a pseduorotaxane complex [**L**·**M**] is established. In this case, both **L** and [**L**·**M**] can react with the stopper **S** to produce a thread and [2]rotaxane. These two products can take part in separate, yet interconnected auto- and crosscatalytic pathways (**Figure 6.11**).

As in the previous chapters, the design of a replicating network integrating rotaxane formation relies on hydrogen-bonding mediated recognition between an amidopyridine unit on linear component **L** and a carboxylic acid positioned on the stopper **S**, for the assembly of the key catalytically active quaternary complex [**R**·**L**·**M**·**S**] (**Figure 6.15a**).

These recognition features are identical to that employed in the formation of an efficient **T1** self-replicating system (**Chapter 2** to **4**) Similarly, macrocycle-binding is likewise driven by hydrogen-bonding recognition between an amide group on the linear component and a complementary recognition units on the macrocycle. In this work, two macrocycles are explored for the formation of a [2]rotaxane, a glycol chain macrocycle **GM**[305–310] and a pyridine spacer macrocycle **PM**.[311]

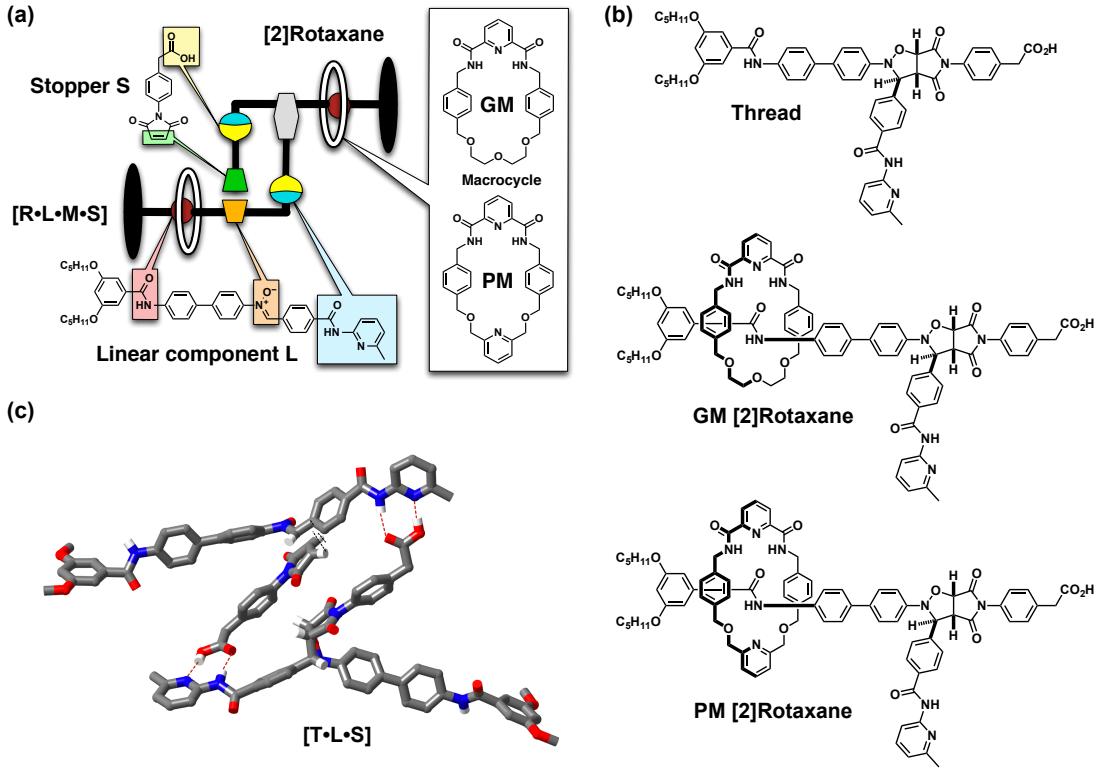


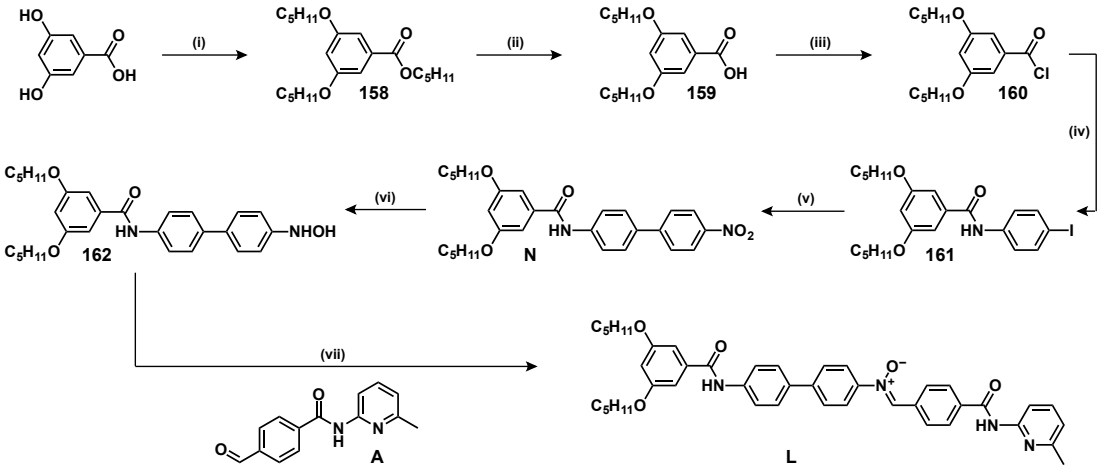
Figure 6.15 (a) Cartoon and chemical structure representation of the components utilised in the design of a self-replicating rotaxane. Amidopyridine (light blue) and carboxylic acid (yellow) recognition sites mediate (red dashed lines) replication of the rotaxane *via* the formation of a quaternary catalytically-active complex [**R**·**L**·**M**·**S**]. Within the complex, the nitrone (orange) and maleimide (green) reactive sites react (black dashed lines) *via* 1,3-dipolar cycloaddition to afford the [2]rotaxane. (b) Chemical structure representation of the thread **T**, GM [2]rotaxane **GM-R** and PM [2]rotaxane **PM-R**. (c) Stick representation of the transition state geometry (RM1, MOPAC 2012) for the formation of thread, illustrating the arrangement of components within a catalytically-active ternary complex [**T**·**L**·**S**] (carbon is represented in grey, nitrogen in blue, oxygen in red and hydrogen in white; most H atoms are omitted for clarity). Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

The irreversible steps leading to the formation of the thread and rotaxane templates exploit 1,3-dipolar cycloaddition reactions between a nitrone (component **L** and [**L**·**M**]) and a maleimide (stopper **S**), resulting in the formation of a thread **T** (**Figure 6.15b**, **T**) and a [2]rotaxane. The reaction of the [**L**·**GM**] or [**L**·**PM**] with **S** locks the macrocycle on the the linear component *via* the stoppering approach, producing a **GM** [2]rotax-

ane (**GM-R**) and **PM** [2]rotaxane (**PM-R**) (**Figure 6.15b**). Following an initial lag period where each thread and rotaxane product is produced predominantly through the slow and unselective bimolecular pathway, the template-mediated catalytic pathway (**Figure 6.15c**) should begin to perform efficiently and with high diastereoselectivity once a sufficient quantity of template has been formed.

6.4.1.1 Synthesis of the linear component

The synthesis of maleimide stopper[c] **S** and aldehyde **A** have been reported previously in **Chapter 3**. Formation of linear component **L** (**Scheme 6.7**) began with alkylation of a commercially available 3,5-dihydroxybenzoic acid with bromopentane, in the presence of a catalytic amount of KI. The intermediate ester **158** was formed in 90% yield, and its deprotection under aqueous basic conditions yielded the alkylated acid **159**, in 80% yield.



Scheme 6.7 Conditions and reagents employed in the synthesis of linear component **L**. (i) KI, K_2CO_3, bromopentane, MeCN/DMF, 92 °C, 16 h, 90%. (ii) *aq.* KOH (10 M), THF, 82 °C, 16 h, 80%. (iii) $SOCl_2$, PhMe, 95 °C, 16 h, quantitative conversion. (iv) 4-iodoaniline, Et_3N, CH_2Cl_2, 0 °C to RT, 16 h, 74%. (v) 4-nitrobenzeneboronic acid, K_3PO_4, $Pd(PPh_3)_4$, EtOH/PhMe/H_2O (1 : 2 : 1), 100 °C, 18 h, 60%. (vi) $NH_2NH_2{\cdot}H_2O$, Rh/C (5% wt), THF, RT, 40 min, quantitative conversion. (vii) $CHCl_3$/EtOH (2 : 1), 4 °C, 4 days, 61%.

The carboxylic acid **159**, now equipped with two solubilising pentyloxy chains, was activated as the corresponding acid chloride **160** using $SOCl_2$ in PhMe, and was reacted immediately with 4-iodoaniline, to give the product **161**, now bearing the amide macrocycle binding site (74% over two steps). In the next step, the iodo precursor **161** was coupled to 4-nitrophenylboronic acid using Suzuki reaction, affording the key intermediate nitro compound **N** (60%). Catalytic amount of rhodium catalyst

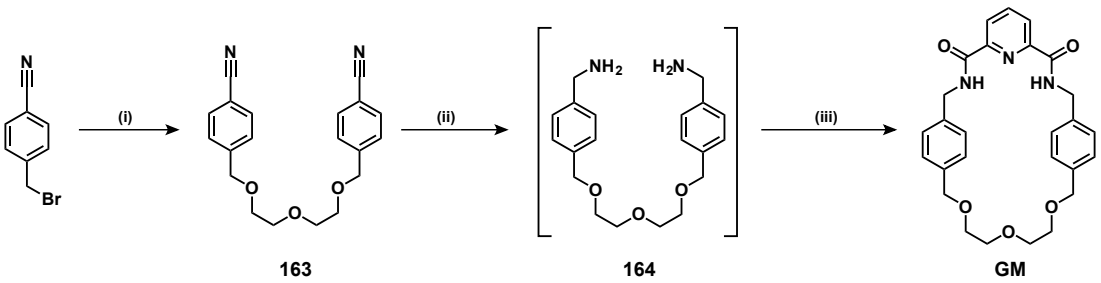
[c]The stoppering group is identical to maleimide **M1** used to form replicator **T1** in **Chapter 3** to **5**. For clarity, because of its role in the formation of the rotaxane, this compound will be referred to as stopper **S** throughout this chapter.

and hydrazine monohydrate were employed to reduce the nitro group in **N** to the corresponding hydroxylamine **162**. Direct condensation of hydroxylamine **162** with aldehyde **A** yielded the desired linear component **L** in 61% yield.

6.4.1.2 Syntheses of macrocycles

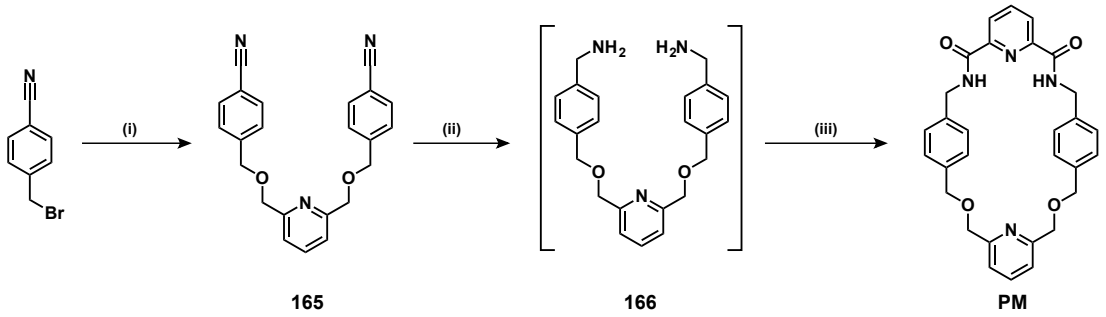
Two structurally-similar macrocycles were exploited in the formation of the self-replicating rotaxane, termed glycol macrocycle **GM** and pyridine macrocycle **PM**. Both the macrocycle **GM** and **PM** have been employed[302,303] in the previous studies examining replicating system based on Model 1. Macrocycle **PM**, bearing a pyridine linker, was expected to provide a stronger association with linear component **L** as a result of the additional pyridine ring with hydrogen-bond acceptor abilities.

Synthesis of macrocycle **GM** (**Scheme 6.8**) commenced with reaction of diethylene glycol and 4-bromobenzonitrile in the presence of NaH (99%). The dinitrile **163**, containing a glycol spacer, was reduced to the corresponding diamine **164** using BH_3·THF complex. The crude product **164** was coupled immediately with 2,6-pyridinedicarbonyl dichloride under dilute reaction conditions, in order to disfavour polymerisation. One equivalent of 4-pyridione was added to the reaction mixture to template the synthesis of the macrocycle. After five days, the target macrocycle **GM** was produced in a very good 38% yield.



Scheme 6.8 Conditions and reagents employed during the synthesis of glycol macrocycle. (i) NaH (60% dispersion in mineral oil), diethylene glycol, THF, 70 °C, 16 hours, 99%. (ii) BH_3·THF, THF, 80 °C, 16 hours, 90%. (iii) 2,6-pyridinedicarbonyl dichloride, 4-pyridione, Et_3N, CH_2Cl_2, RT, 5 days, 38%.

Macrocycle **PM** was prepared (**Scheme 6.9**) using a similar synthetic route. In the first step, reaction between 4-bromobenzonitrile and 2,6-pyridine dimethanol in the presence of NaH afforded a dinitrile compound **165** (96%). The dinitrile compound **165**, bearing a pyridine spacer, was again reduced to the corresponding diamine **166** using BH_3·THF complex (90%). The crude diamine **166** was coupled immediately with 2,6-pyridinedicarbonyl dichloride under dilute reaction conditions in order to discourage polymerisation. After 13 days, the target macrocycle **PM** was produced in 11% yield.



Scheme 6.9 Conditions and reagents employed during the synthesis of pyridine macrocycle. (i) NaH (60% dispersion in mineral oil), 2,6-pyridine dimethanol, THF, 70 °C, 2 days, 96%. (ii) BH_3·THF, THF, 70 °C, 18 hours, 90%. (iii) 2,4-pyridinedicarbonyl dichloride), Et_3N, CH_2Cl_2, RT, 13 days, 11%.

6.4.2 Pseudorotaxane binding studies

As the first step in the investigation of this rotaxane forming network, the binding between each macrocycle and a simpler analogue, **N**, of the full linear component **L**, was investigated. The model compound **N**, bearing a NO_2 group, does not possess the amidopyridine recognition site or the nitrone reactive unit, allowing the binding of **GM** and **PM** with the target amide macrocycle binding site to be examined in isolation of any other recognition processes.

To this end, a 10 mM solution containing both **N** and **GM** was prepared in $CDCl_3$ and examined by 500.1 MHz ^{1}H NMR spectroscopy. At room temperature, only broad resonances corresponding to the pseudorotaxane [**N**·**GM**] were observed. By reducing the temperature to −40 °C, the exchange between the bound and unbound state entered a slow exchange regime (**Figure 6.16b**) on the NMR chemical shift time-scale, allowing direct observation of the resonances arising from **N** (**Figure 6.16a**) and **GM** (**Figure 6.16c**) as well as the pseudorotaxane complex [**N**·**GM**].

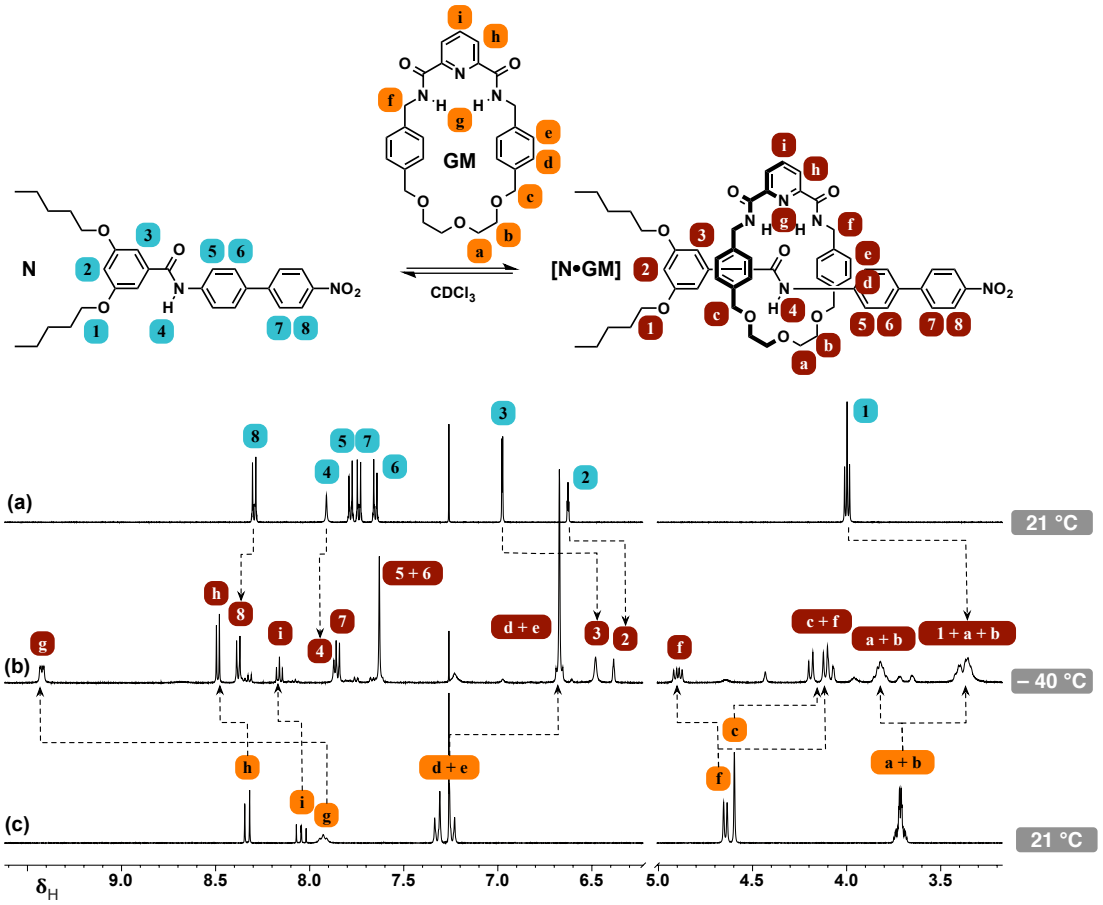


Figure 6.16 Partial ^{1}H NMR (500.1 MHz, $CDCl_3$) spectra of **(a)** nitro compound **N** (21 °C), **(b)** an equimolar mixture of nitro compound **N** and **GM** (10 mM, −40 °C) and **(c) GM** (21 °C). The dashed lines connect the resonances arising from specific protons in the uncomplexed **N** and **GM** to the corresponding resonances in the bound pseudorotaxane complex. Unassigned resonances in **(b)** correspond to the unbound species. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

Directly indicative of pseudorotaxane complex formation was a significant downfield shift of the resonance corresponding to the macrocycle NH protons (**Figure 6.16**, H^g, +1.49 ppm). This change in the chemical shift of the macrocycle NH resonances within the bound complex (9.42 ppm) suggest that these protons are taking part in hydrogen-bonding mediated interaction with the carbonyl of the amide macrocycle-binding site in **N**. The phenylene proton resonances associated with H^d and H^e were shifted upfield (−0.66 and −0.56 ppm, respectively), behaviour typical for protons located in the shielding zone of an aromatic ring. Similarly, the resonances corresponding to H^a, H^b and H^c on the macrocycle have become magnetically-inequivalent upon formation of the pseudorotaxane, as a result of an end-to-end asymmetry imposed on the **GM** by the nitro compound **N**. These changes are consistent with macrocycle being bound at the desired amide-binding site within the structure of **N**. The sharp resonances observed in

the low temperature spectrum of this complex also allowed the K_a for this interaction to be determined as 3750 M^{-1} using a single-point method.

Similar analysis was undertaken for the formation of pseudorotaxane with macrocycle **PM**. To this end, an equimolar solution of **N** and **PM** (10 mM) was prepared in $CDCl_3$. ^{1}H NMR spectra (500.1 MHz) were periodically acquired until the pseudorotaxane formation was in slow exchange mode (−50 °C) on the NMR time-scale, and sharp resonances for the bound and free species could be observed (**Figure 6.17b**). Similar chemical shift changes, indicative of pseudorotaxane formation, were observed for the resonances arising from macrocycle **PM** and **N** within [**N**·**PM**] compared to the corresponding free species, as was determined for the [**N**·**GM**] pseudorotaxane. From this data, it was again possible to calculate the K_a value for the pseudorotaxane [**N**·**PM**] formation at −50 °C as 5100 M^{-1}. The higher association constant determined for the association of **PM** with **N** reflects the additional pyridine recognition unit present in **PM**.

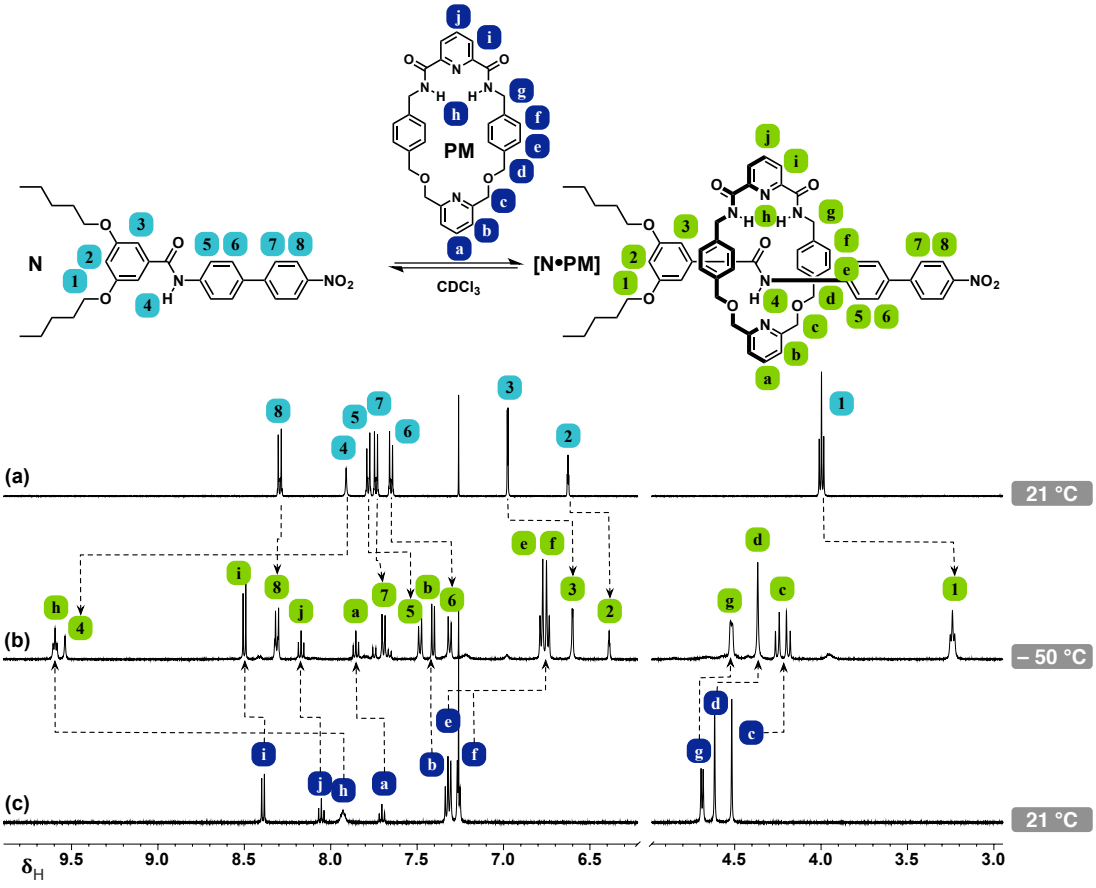


Figure 6.17 Partial ^{1}H NMR (500.1 MHz, $CDCl_3$) spectra of **(a)** nitro compound **N** (21 °C), **(b)** an equimolar mixture of nitro compound **N** and **PM** (10 mM, −50 °C) and **(c)** **PM** (21 °C). The dashed lines connect the resonances arising from specific protons in the free, uncomplexed **N** and **PM** to the corresponding resonances in the bound pseudorotaxane complex. Unassigned resonances in **(b)** correspond to the unbound species. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

In addition to determination of the association constant, it was also possible to confirm the location of the **PM** macrocycle on the target binding site in **N** using low-temperature (−50 °C) ^{1}H–^{1}H rotating frame Overhauser effect spectroscopy (ROESY) (500.1 MHz) (**Figure 6.18**). The ^{1}H–^{1}H showed rOe cross peaks between the resonances associated with the NH protons on the macrocycle (H^{a}) and the resonances corresponding to the phenyl rings on each side of the amide group in **N** (H^{1} and H^{2}). Furthermore, cross peaks were observed between the resonances arising from the —CH_2 in **PM** (H^{b}) and the resonances corresponding to the NH in compound **N** (H^{3}). Despite the sharp resonances observed for all species at −40 °C, the ^{1}H–^{1}H ROESY NMR analysis of [**N**·**GM**] did not provide data of sufficient quality for unambiguous confirmation of the macrocycle position within the [**N**·**GM**] pseudorotaxane.

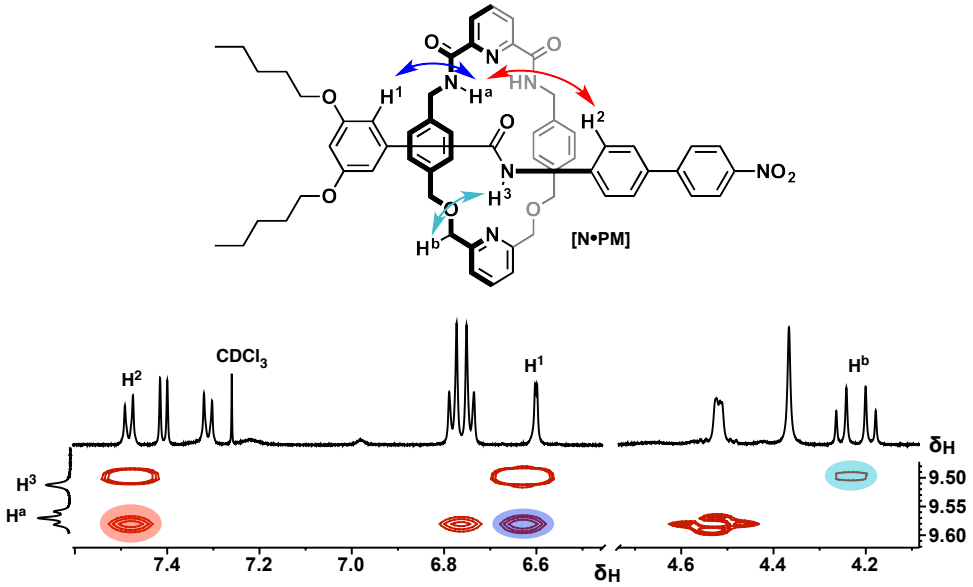


Figure 6.18 Partial ^{1}H–^{1}H ROESY correlation of pseudorotaxane [**N**·**PM**] (500.1 MHz, $CDCl_3$, RT, 10 mM, −50 °C, 200 ms mixing time). Highlighted rOe cross peaks represent the through-space interactions between the model nitro compound **N** and the macrocycle **PM**. Unassigned cross peaks represent interactions within the individual components.

Having established the association between each macrocycle and the model compound **N**, containing only the target amide macrocycle-binding site, the next step was to probe the association of linear component **L** with **GM** and **PM**. The linear component is equipped with an amidopyridine recognition site and a reactive nitrone site, to which the macrocycles might potentially also bind—the negatively charged oxygen atom of the nitrone moiety has the potential[312,313] to be a good H-bond acceptor.

An equimolar solution of linear component **L** and glycol macrocycle **GM** (10 mM) was prepared in $CDCl_3$. ^{1}H NMR spectra (500.1 MHz) were acquired at a series of temperatures between −60 °C and 21 °C. The temperature range available through the use of $CDCl_3$ as a solvent, however, was not sufficient to achieve slow exchange regime

on the NMR chemical shift time-scale and sharp resonances for the bound and free species could not be observed, even at −60 °C (**Figure 6.19**). Calculation of the K_a value for the pseudorotaxane formation was therefore not possible from these data.

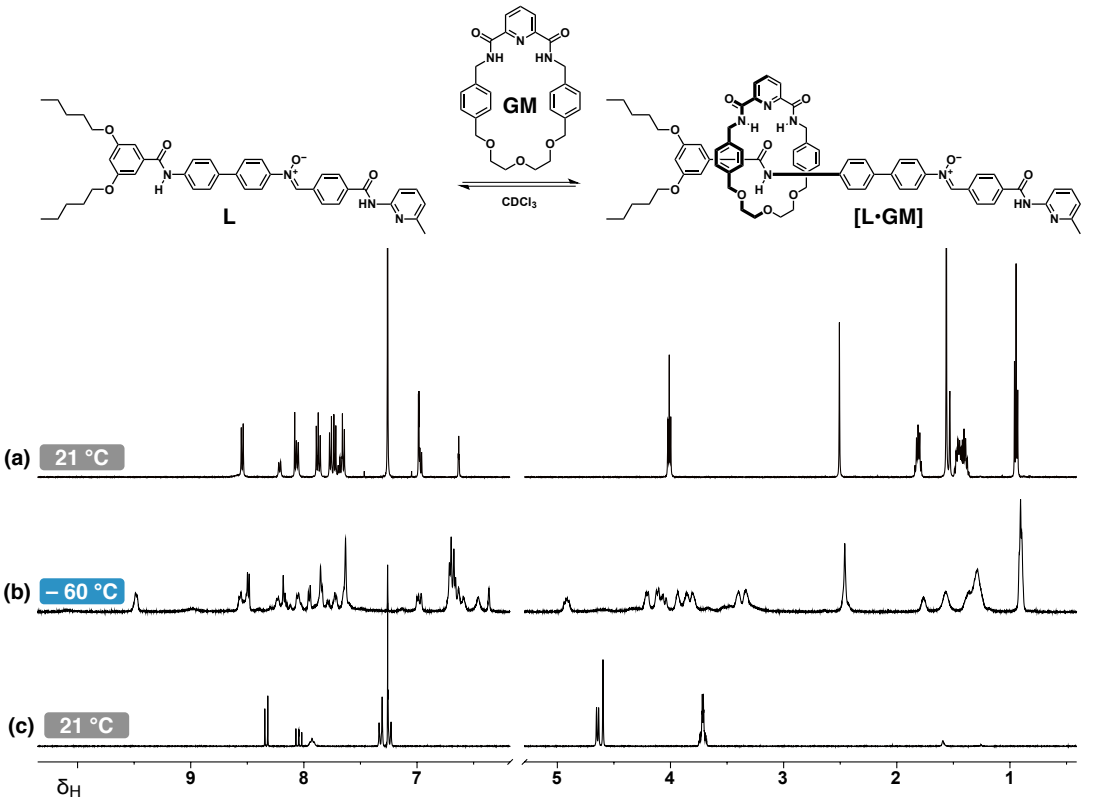


Figure 6.19 Partial ^{1}H NMR (500.1 MHz, $CDCl_3$) of **(a)** linear component **L** (21 °C), **(b)** an equimolar mixture of linear component **L** and **GM** (−60 °C), (c) **GM** (21 °C).

An identical procedure was employed for the analysis of an equimolar solution of the linear component and **PM** in $CDCl_3$ (10 mM, $CDCl_3$). ^{1}H NMR spectra (500.1 MHz were acquired at a series of temperatures −30 °C and 21 °C, however, the slow exchange on the NMR chemical shift time-scale was not achieved. **Figure 6.20** shows the ^{1}H NMR (500.1 MHz) spectrum recorded at −30 °C, temperature at which the resonances arising from the bound and unbound components were observed to be the sharpest. Throughout the cooling processes, the resonances for both the pseudorotaxane and the free species were observed to go through a maximum sharpness at −30 °C, which broadened again upon further cooling, making the calculation of the K_a for the formation of the pseudorotaxane impossible from the available data.

While it was not possible to determine the K_a values for the formation of [**L**·**GM**] and [**L**·**PM**] complexes, the association is clearly evidenced by the appearance of new, albeit broad resonances arising from the complexed species, indicating that the system should be suitable for rotaxane assembly.

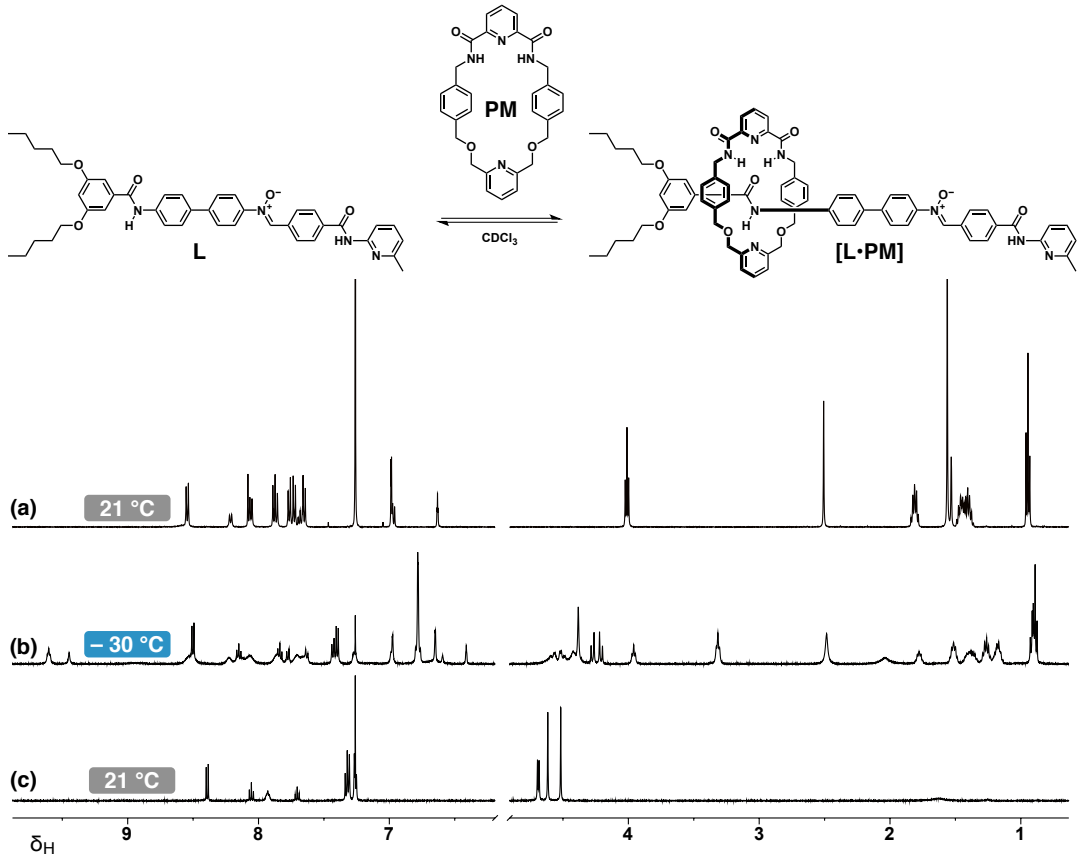
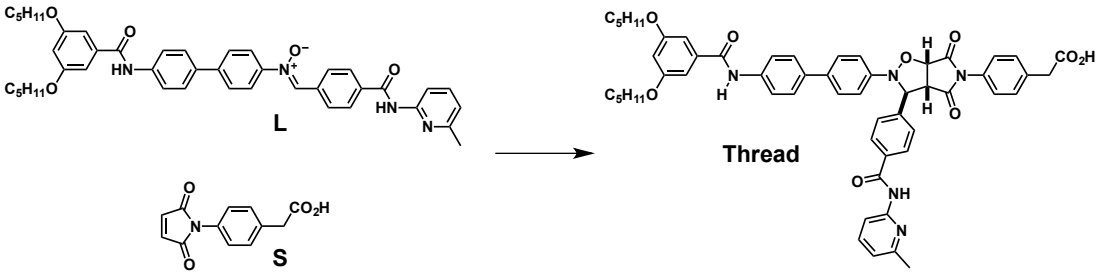


Figure 6.20 Partial ^{1}H NMR spectrum (500.1 MHz, $CDCl_3$) of **(a)** linear component **L** (21 °C), **(b)** equimolar mixture of linear component **L** and **PM** (−30 °C), **(c) PM** (21 °C).

6.4.3 Synthesis and characterisation of the thread and rotaxanes

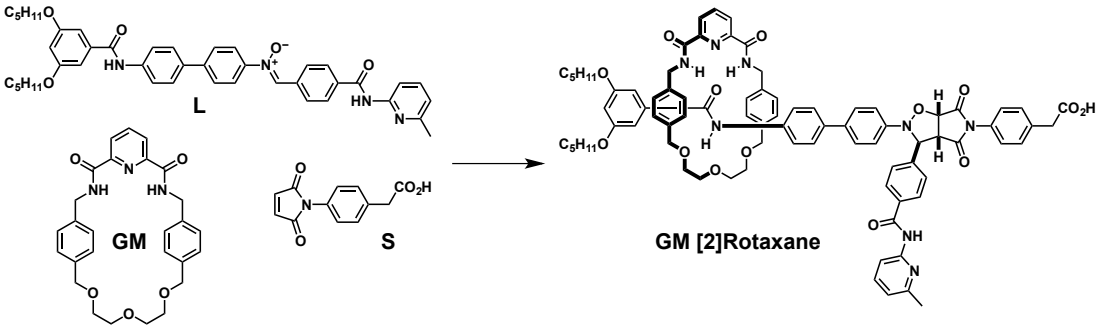
Having established the promising nature of the designed system for rotaxane formation, the next step was to confirm that the envisaged thread and both rotaxanes, can, in fact, be synthesised. In order to confirm that their formation is possible, and to prepare sufficient quantities of preformed templates for use in instructed kinetic experiments, the thread and two rotaxanes were synthesised on a preparative scale from the individual components. Thread **T** was prepared (**Scheme 6.10**) by reaction of linear component **L** and stopper **S** in $CDCl_3$. After two days at 4 °C, ^{1}H NMR spectroscopy confirmed complete disappearance of the starting materials. Removal of the solvent under a flow of compressed air produced the crude thread **T**. Isolation[d] of pure sample of thread (45%), in sufficient quantity for characterisation and for use in kinetic analyses, was achieved using reverse phase chromatography (10 g ISOLUTE C18 column).

[d]Analysis of the crude samples of thread and both rotaxanes by NMR spectroscopy revealed full conversion to products—the yields determined for **T**, **GM-R** and **PM-R** by NMR spectroscopy were significantly higher than the yields determined following isolation of these components after purification by reverse phase column chromatography.

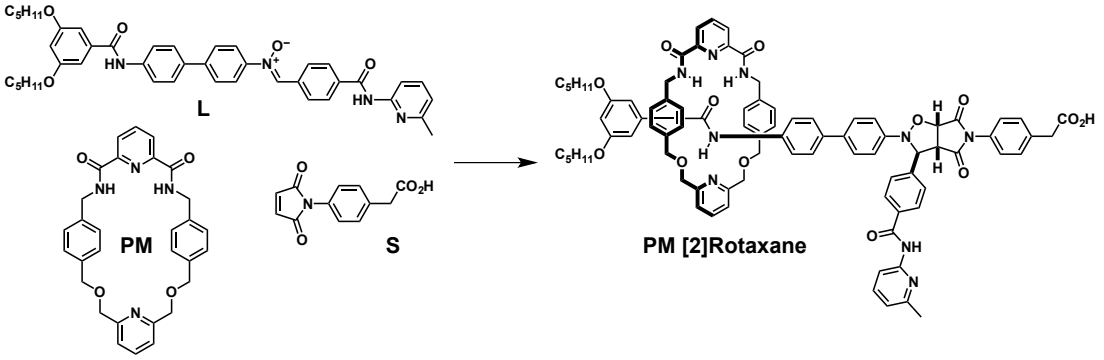


Scheme 6.10 Conditions and reagents employed in the synthesis of thread: $CDCl_3$, 4 °C, 2 days, 45% isolated yield.

The **GM-R** and **PM-R** rotaxanes were prepared by reaction of linear component **L** with stopper **S** in the presence of either **GM** (**Scheme 6.11**) or **PM** (**Scheme 6.12**), in $CDCl_3$ at low temperatures. After consumption of the starting material was confirmed by ^{1}H NMR spectroscopy, the reaction mixtures were concentrated under a flow of compressed air. The crude samples of each rotaxane were purified by reverse phase chromatography (20 g ISOLUTE C18 columns), providing a sufficient quantity of each of the rotaxane *trans* template for characterisation and kinetic experiments (15% and 16% yield for **GM-R** and **PM-R**, respectively).



Scheme 6.11 Conditions and reagents employed in the synthesis of GM [2]Rotaxane: $CDCl_3$, −20 °C → 4 °C, 18 days, 15% isolated yield.



Scheme 6.12 Conditions and reagents employed in the synthesis of PM [2]rotaxane: $CDCl_3$, −20 °C → 4 °C, 9 days, 16% isolated yield.

The identity of both rotaxanes was confirmed using matrix-assisted laser desorption/ionisation (MALDI) time-of-flight mass spectrometry, ^{1}H and ^{13}C NMR spectroscopy and combination of 2D NMR experiments. A single crystal, suitable for X-ray crystallographic analysis, was obtained for both rotaxanes. The X-ray analyses confirmed (**Figure 6.21**) the expected location of each macrocycle on the amide binding site on the threading component, mediated by hydrogen bonding interactions. The crystallographic analysis also confirmed the predicted *trans* diastereochemistry of the isoxazolidine ring formed in the 1,3-dipolar cycloaddition reaction, and the 'open' conformation of the recognition sites associated with this conformation, necessary for template-directed replication.

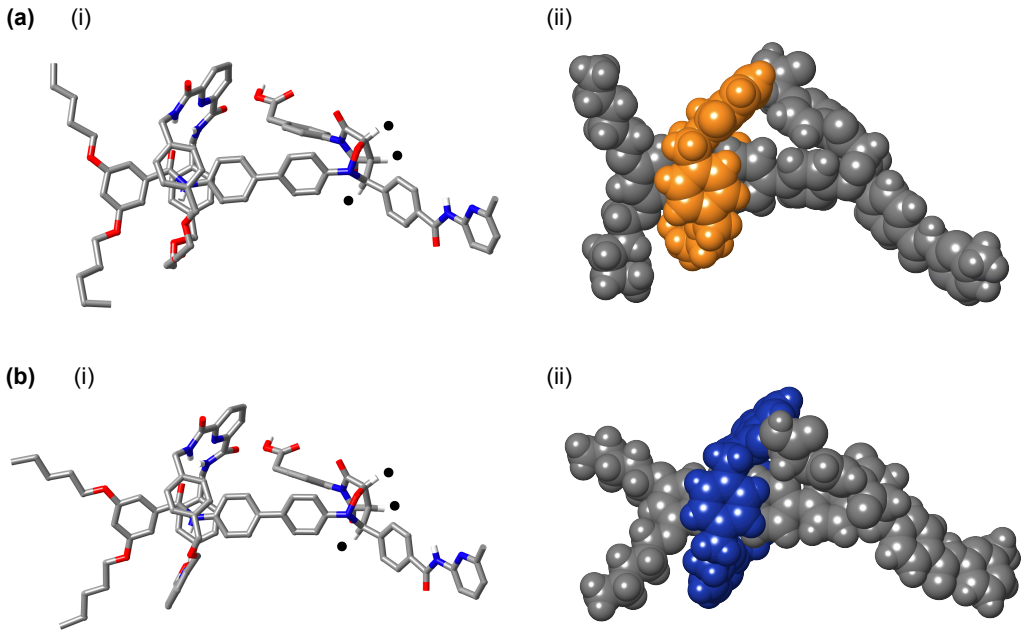


Figure 6.21 Stick (i) and space-fill (ii) representations of the X-ray crystal structures obtained for **(a)** GM [2]rotaxane and **(b)** PM [2]rotaxane by X-ray crystallographic analysis of single crystals. In the stick representations, C atoms are coloured grey, O atoms in red, N atoms in blue and H atoms in white. Majority of hydrogen atoms are omitted for clarity. The black dots in (i) highlight the *trans* stereochemistry of the protons on the isoxazolidine ring in **GM-R** and **PM-R**. The colouring in the space-fill representations shows the threading component in grey, **GM** in orange and **PM** in blue. Data collection and X-ray diffraction analysis were performed by Dr. D. B. Cordes and Professor A. M. Z. Slawin at the University of St Andrews. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

The expected location of each macrocycle within the [2]rotaxane structure at room temperature was confirmed by two-dimensional ^{1}H–^{1}H ROESY experiments. **Figure 6.22a** shows the rOe cross peaks observed in the 2D ROESY NMR spectrum of **GM-R**.

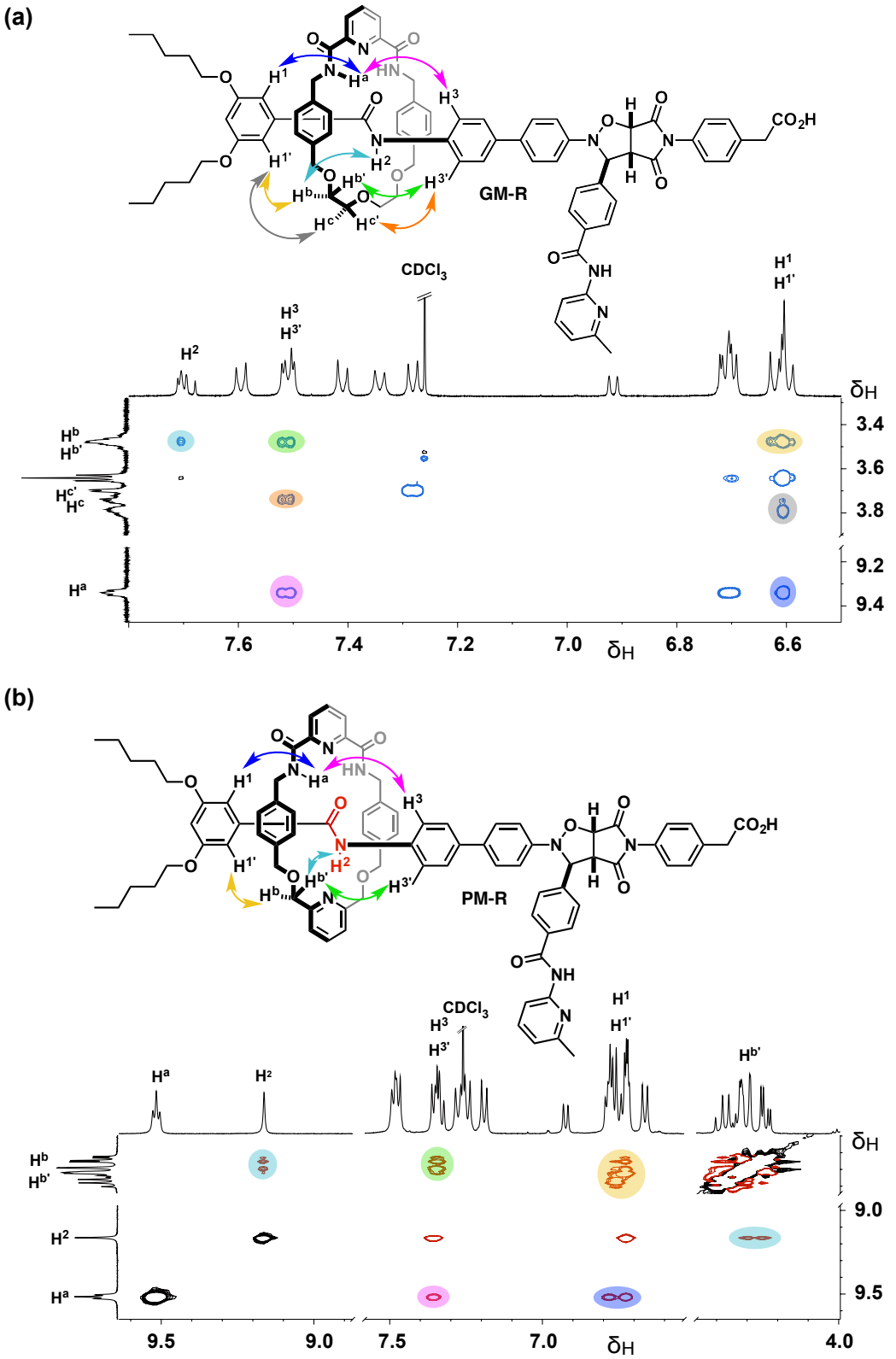


Figure 6.22 Partial ^{1}H–^{1}H ROESY correlation of **(a)** GM [2]rotaxane (500.1 MHz, $CDCl_3$, RT, 500 ms mixing time) and **(b)** PM [2]rotaxane $CDCl_3$, RT, 200 ms mixing time). Highlighted rOe cross peaks represent the through-space interactions between the threading component and the macrocycle. Unassigned cross peaks represent intra-component interactions. Figure adapted with permission from Ref. 301. Copyright 2015 American Chemical Society.

The highlighted cross peaks particularly indicative of the macrocycle binding at the target location are those between resonances arising from the macrocycle H^a, H^b, and H^c and the resonances arising from H^1, H^2, and H^3 protons on the threaded linear component. Similarly, the partial 2D ROESY spectrum of **PM-R** (**Figure 6.22b**) confirmed that macrocycle **PM** is bound at the desired amide binding site. This association was particularly evident from the cross peaks between the resonances associated with H^a and H^b of **PM** and H^1, H^2 and H^3 on the threading component.

6.4.4 Kinetic analyses

Having synthesised sufficient quantities of **T**, **GM-R** and **PM-R** and undertaken their full characterisation, the next step was to carry out a comprehensive kinetic analysis, designed to assess the capacity of these three species to self-replicate and to take part in crosscatalytic pathways. Ultimately, undertaking of both recognition-disabled and recognition-enabled experiments will ascertain which catalytic pathways are active in each rotaxane–thread replicating system, as well as their efficiency.

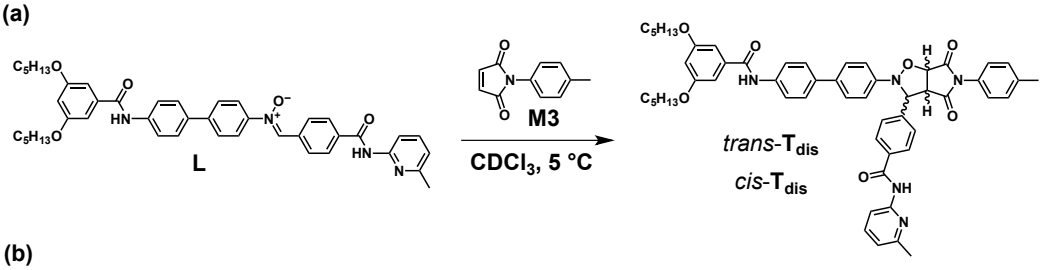
6.4.4.1 Recognition-disabled kinetic experiments

As the first step in the kinetic analyses of the thread and rotaxane systems, formation of the thread and both rotaxane components was examined in recognition-disabled kinetic experiments. These experiments examined the formation of thread and each rotaxane in a scenario where the maleimide —COOH recognition site is substituted with a methyl group—maleimide **M3**). The recognition-disabled experiments were performed as a single time-point NMR experiments, where the required reaction components were mixed together (10 mM), transferred to a 5 mm NMR tube and placed in a temperature controlled water bath, equilibrated at 5 °C. Composition of each sample was analysed by ^{1}H NMR (500.1 MHz) spectroscopy after *ca.* 18 hours. **Figure 6.23a** shows the components and typical reaction conditions employed in the recognition-disabled kinetic experiments, as illustrated with the reaction of linear component **L** with **M3**, to afford the recognition-disabled *trans*-$\mathbf{T_{dis}}$ and *cis*-$\mathbf{T_{dis}}$.

Results of the recognition-disabled NMR experiments (**Figure 6.23b**) show that in the absence of one of the recognition sites required for self-replication, the 1,3-dipolar cycloaddition reactions proceed with very low conversion even after 18 hours, and low diastereoselectivity in the thread as well as the rotaxane forming experiments (on average, [*trans*]/[*cis*] ratio of 4). Furthermore, selectivity for the thread over the rotaxane ([thread]/[rotaxane]) also remained low at a ratio of *ca.* 1.7 and 1.2 in the GM [2]rotaxane and PM [2]rotaxane experiments, respectively. The slightly higher

concentration of the recognition-disabled **PM-R$_{dis}$** is in agreement with the higher K_a for the formation of [**N**·**PM**] than the corresponding pseudorotaxane with **GM**.

(a)



(b)

NMR Experiment	Thread$_{dis}$ / mM		Rotaxane$_{dis}$ / mM		Overall conversion / %
	trans	*cis*	*trans*	*cis*	
Thread	1.21	0.28	–	–	15
GM-R	0.55	0.12	0.31	0.09	11
PM-R	0.50	0.11	0.41	0.11	11

Figure 6.23 **(b)** Example recognition-disabled kinetic experiment examining the reaction between linear component **L** and maleimide lacking the carboxylic acid recognition site **M3** ([**L**] = [**M3**] = 10 mM, $CDCl_3$, 5 °C). **(b)** Product distribution and conversion determined in recognition-disabled kinetic experiments examining the formation thread ([**L**] = [**M3**] = 10 mM), GM [2]rotaxane ([**L**] = [**M3**] = [**GM**] = 10 mM) and PM [2]rotaxane ([**L**] = [**M3**] = [**PM**] = 10 mM) equipped with only a single recognition site, as determined by ^{1}H NMR spectroscopy (500.1 MHz) after *ca.* 18 hours ($CDCl_3$, 5 °C, 2,4-dinitrotoluene as internal standard). As a result of overlapping resonances in the **PM-R** recognition-disabled experiment, the formation of *cis* products is determined based on an estimated 1 : 1 thread:rotaxane product distribution.

6.4.4.2 Kinetic analyses of thread

Having established that thread, GM [2]Rotaxane and PM [2]Rotaxane formation is slow and exhibits poor diastereoselectivity in the absence of recognition elements, the formation of the thread from linear component **L** and stopper **S** was examined. Initially, the formation of thread **T** was examined in the absence of any added pre-formed template. An equimolar mixture containing the reaction components ([**L**] = [**S**] = 10 mM) was prepared in $CDCl_3$ and the reaction progress at 5 °C was monitored by ^{1}H NMR (500.1 MHz) spectroscopy every 15 minutes over four hours. The reaction profile (**Figure 6.24a**, black) for this reaction was determined by examining the resonances for the three protons corresponding to the *trans* isoxazolidine ring of the cycloadduct relative to the resonances arising from the protons in the CH_3 group of 2,4-dinitrotoluene as an internal standard. Following an initial lag period, typical for self-replicating systems, the *trans*-**T** replicated very efficiently, exhibiting a clear sigmoidal reaction profile, reaching a concentration of 9.1 mM after four hours (91% conversion). The rate (**Figure 6.24b**, black) for this reaction (d[**T**]/dt) was determined by computing the first derivative of a seventh-order polynomial fitted to the concentration *vs* time data. The

trans diastereoisomer was formed exclusively, with the rate maximum observed after 63 min (4.1 mMh^{-1}). Next, the formation of thread *via* the autocatalytic pathway was examined by undertaking the same reaction in the presence of instructing preformed thread template. The required starting components ([**L**] = [**S**] = 10 mM) and preformed thread (10 mol%) were dissolved in $CDCl_3$ and the reaction progress was followed by ^{1}H NMR spectroscopy as described previously. Addition of preformed thread at $t = 0$ resulted in the disappearance of the lag period (**Figure 6.24a**, grey), allowing the system to reach its maximum rate (6.0 mMh^{-1}) at the beginning of the reaction (**Figure 6.24b**, grey), thereby confirming the successful design of the self-replicating system at the core of the thread–rotaxane network. With this information in hand, thread formation was examined in the presence of 10 mol% of preformed **GM-R** (**Figure 6.24a**, orange) and **PM-R** (**Figure 6.24a**, blue).

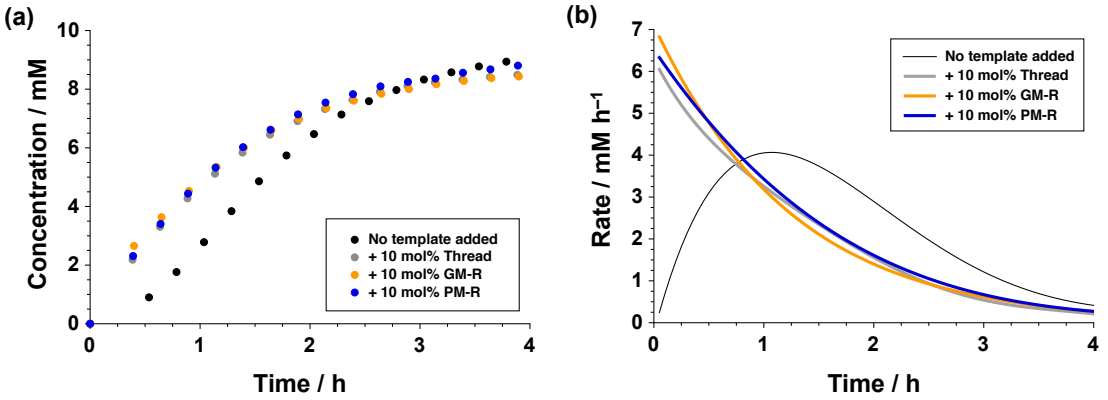


Figure 6.24 **(a)** Concentration and **(b)** rate *vs* time profiles for the formation of **T** from linear component **L** and stopper **S** in the absence of added template (black) and in the presence of 10 mol% preformed **T** (grey), 10 mol% preformed **GM-R** (orange) and 10 mol% preformed **PM-R** (blue) as determined by ^{1}H NMR spectroscopy (500.1 MHz, $CDCl_3$, 5 °C, all components at 10 mM). Concentrations were determined relative to 2,4-dinitrotoluene as an internal standard. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

The reaction profiles for the thread kinetic experiments instructed with preformed **GM-R** and **PM-R** closely mirrored the formation of thread determined in the presence of preformed thread template. In the rotaxane instructed experiments, a clear disappearance of the lag period is observed, together with a shift in the time of the maximum reaction rate (**Figure 6.24b**). Specifically, maximum reaction rate of 6.8 and 6.3 mMh^{-1} is observed for **GM-R** and **PM-R**, respectively at $t = 0$. Comparison of the uninstructed experiment with the three template doped experiments revealed that thread is indeed self-replicating, and the two rotaxanes are capable of crosscatalysing the formation of the thread with efficiency equal to that of thread itself.

6.4.4.3 Kinetic analyses of rotaxanes

Having undertaken the full analysis of the thread kinetic pathways, the next step was to examine the individual, uninstructed formation of **GM-R** (**Figure 6.25**) and **PM-R** (**Figure 6.26**). The progress of the rotaxane forming kinetic experiments was followed by 500.1 MHz ^{1}H NMR spectroscopy, which permitted deconvolution of the individual proton resonances arising from the isoxazolidine ring of each thread and rotaxane cycloadduct (for details, see **Chapter 9**).

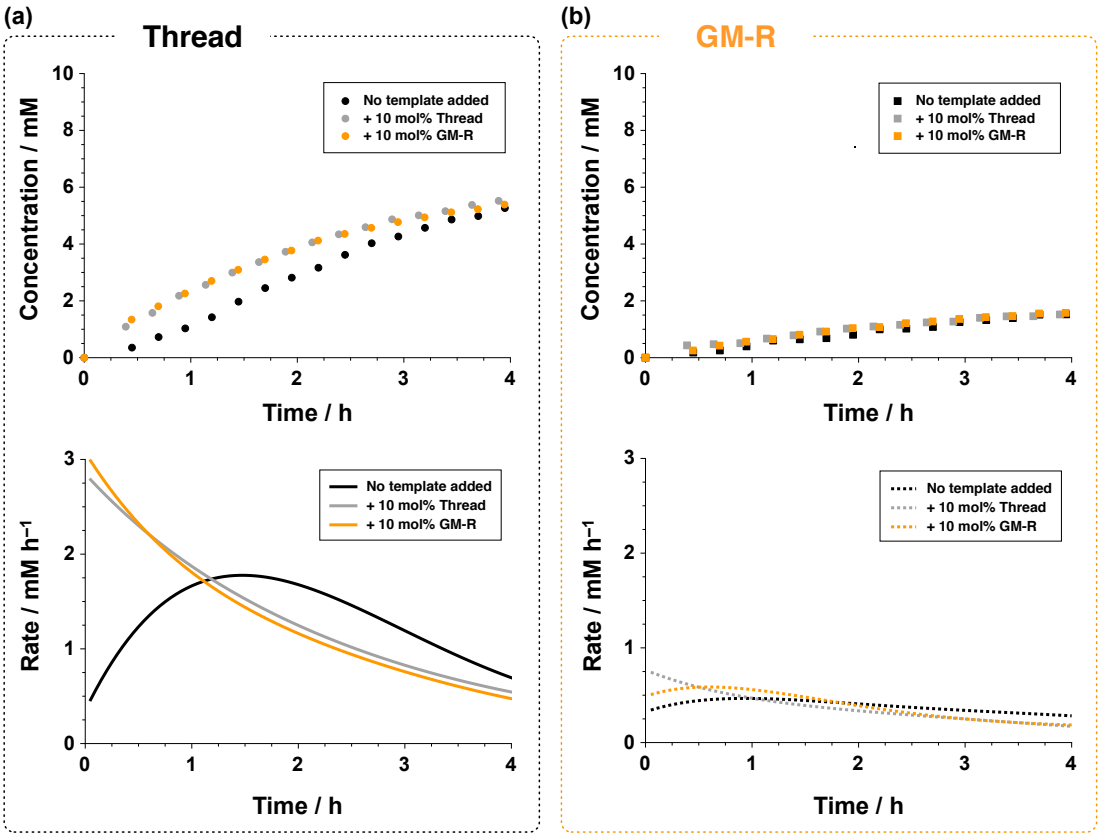


Figure 6.25 Concentration (top) and rate (bottom) profiles for the formation of **(a) T** (circles) and **(b) GM-R** (squares) in the kinetic experiment examining reaction of linear component **L** and stopper **S** in the presence of macrocycle **GM**. The kinetic experiments were undertaken in the absence of added template (black) and in the presence of 10 mol% preformed **T** (grey) and 10 mol% preformed **GM-R** (orange), as determined by ^{1}H NMR spectroscopy (500.1 MHz, $CDCl_3$, 5 °C, all components at 10 mM). Concentrations were determined relative to 2,4-dinitrotoluene as an internal standard. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

The concentration-time profiles for the **GM-R** kinetic experiments (**Figure 6.25**, top), examining the reaction of **L** and **S** in the presence of one equivalent of **GM**, show that both **T** and **GM-R** were formed with high diastereoselectivity for the *trans* cycloadducts. Within the uninstructed **GM-R** kinetic experiment, the formation of **T** proceeded less efficiently, reaching concentration of only 5.2 mM after four hours.

The catalytic pathways leading to the formation of **GM-R** performed less efficiently, producing the rotaxane at a concentration of 1.5 mM after four hours. When the same experiment was performed in the presence of 10 mol% of preformed instructing thread template, formation of both products proceeded more efficiently (**Figure 6.25**, grey). The reaction profiles observed in the rotaxane kinetic experiments instructed with 10 mol% **GM-R** (**Figure 6.25**, orange) mirrored the corresponding profiles obtained in the thread instructed experiments very closely, showing a nearly identical ratio of [**R**]/[**T**] (*ca.* 3) and a shortened lag period across each rotaxane time course, as evidenced by the shift in maximum rate for the formation of thread and **GM-R** to an earlier time point (**Figure 6.25**, bottom) in both thread and **GM-R** instructed kinetic experiments.

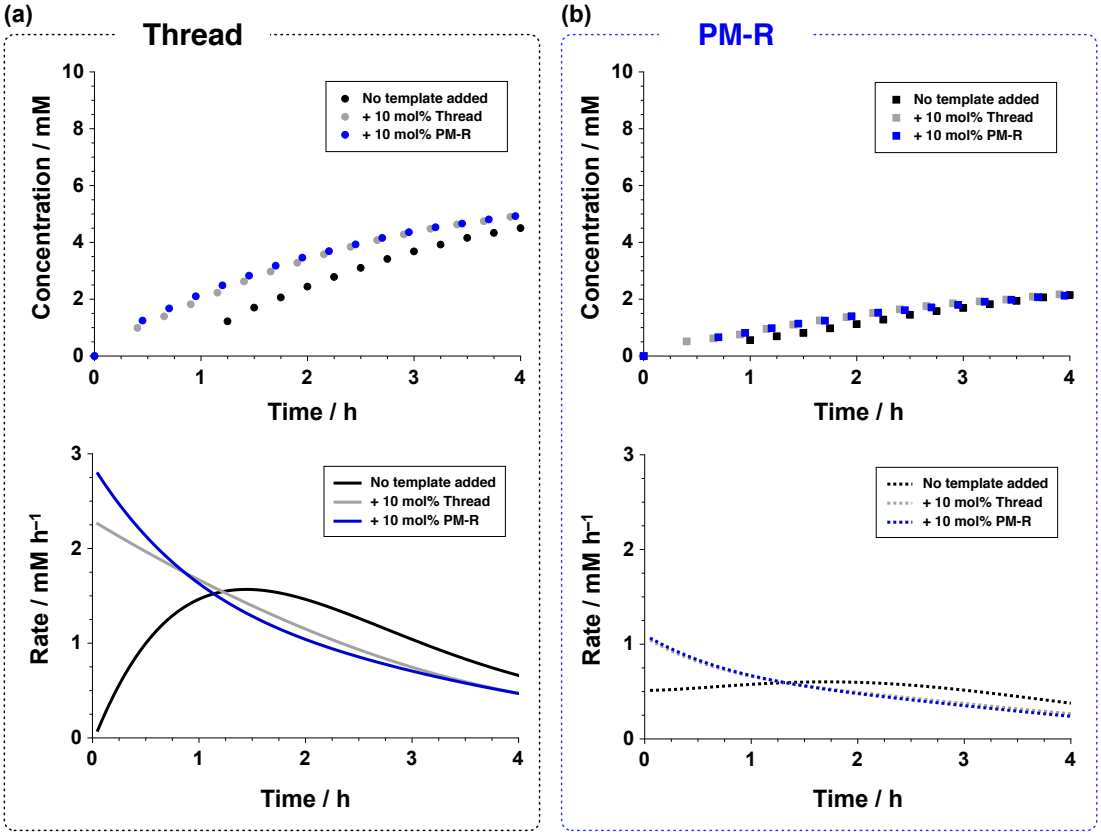


Figure 6.26 Concentration (top) and rate (bottom) profiles for the formation of **(a) T** (circles) and **(b) PM-R** (squares) in the kinetic experiment examining reaction of linear component **L** and stopper **S** in the presence of macrocycle **PM**. The kinetic experiments were undertaken in the absence of added template (black) and in the presence of 10 mol% preformed **T** (grey) and 10 mol% preformed **PM-R** (blue), as determined by ^{1}H NMR spectroscopy (500.1 MHz, $CDCl_3$, 5 °C, all components at 10 mM). Concentrations were determined relative to 2,4-dinitrotoluene as an internal standard. Figure reprinted with permission from Ref. 301. Copyright 2015 American Chemical Society.

Having examined the formation of **GM-R**, a similar procedure was employed for kinetic analysis of the **PM-R** system. In the first instance, formation of thread and rotaxane from **L**, **S** and one equivalent of **PM** was examined in the absence of

preformed template. As expected based on the higher K_a for the formation of the pseudorotaxane [**L**·**PM**] than [**L**·**GM**], the **PM-R** was formed more efficiently than **GM-R** under identical reaction conditions. Nevertheless, the synthesis of this rotaxane product, while similarly diastereoselective for the *trans* cycloadduct, reached lower conversion than the thread again. The **PM-R** reached a concentration of 2.2 mM after four hours, whereas the thread was formed at concentration of 4.5 mM after four hours. Examination of the **PM-R** system in the presence of 10 mol% of preformed instructing (i) thread (**Figure 6.26**, grey) and (ii) **PM-R** template (**Figure 6.26**, blue) showed a decrease in the length of the lag period for both the thread and the rotaxane component, as well as a shift in the maximum rate for the formation of thread and rotaxane to an earlier time point (**Figure 6.26**, bottom) in both template instructed kinetic experiments. As observed for **GM-R** system, the reaction profiles from the two template instructed experiments matched each other extremely closely, exhibiting a nearly identical ratio of [**R**]/[**T**] (*ca.* 2).

The outcome of these native (template-free) and instructed rotaxane kinetic experiments confirmed that both auto- and crosscatalytic cycles are operating in both rotaxane system with equal efficiency. Satisfyingly, results of these kinetic analyses confirmed that the orthogonal recognition processes enable formation of a replicating network where the thread and the rotaxane are matched in catalytic efficiencies as templates, in both auto- and crosscatalytic cycles. Nevertheless, the ratios of [**R**]/[**T**] determined in our kinetic experiments showed a noticeable preference for the thread over the rotaxane in both replicating systems. The relatively low [**R**]/[**T**] ratio indicates that while the preformed rotaxane templates have the capacity to catalyse the formation of both thread and rotaxane products very effectively, the formation of the rotaxane itself proceeded less efficiently than envisaged based on the low temperature pseudorotaxane experiments.

6.4.5 Kinetic fitting and simulations

One plausible explanation for the low efficiency of rotaxane formation compared to that of thread, and thus the corresponding low [**R**]/[**T**] ratio is the insufficient proportion of each reactive pseudorotaxane complex $[\mathbf{L}\cdot\mathbf{M}]_{amide}$ (**Figure 6.27a**), containing the macrocycle at the desired amide binding site, as a result of the nitrone reactive site and amidopyridine recognition unit also associating with the macrocycles. The desired reactive pseudorotaxane complex $[\mathbf{L}\cdot\mathbf{M}]_{amide}$ has the capacity to react with stopper **S** through a bimolecular reaction ($k_1 = k_{bi}$) and, in the presence of template, also through a template-mediated pseudo-unimolecular reaction ($k_2 = k_{uni}$). By contrast, macrocycle bound at the nitrone reactive site results in formation of a pseudorotaxane $[\mathbf{L}\cdot\mathbf{M}]_{nitrone}$

(**Figure 6.27b**) that sequesters **L** and **M** components in a co-conformation that prevents any reaction.

The association of a macrocycle with the amidopyridine recognition site, required for recognition-mediated self-replication, affords a pseudorotaxane complex $[\mathbf{L}{\cdot}\mathbf{M}]_{\text{amidopyridine}}$ (**Figure 6.27c**) that can only react through the bimolecular pathway ($k_4 = k_{\text{bi}}$). The reaction of pseudorotaxane $[\mathbf{L}{\cdot}\mathbf{M}]_{\text{amidopyridine}}$ with stopper **S** results in the slow formation of thread, with macrocycle bound at the 6-amidopyridine recognition site. Ultimately, the presence of these two additional sites on the linear component, capable of associating with the macrocycle, decreases the proportion of the desired pseudorotaxane complex $[\mathbf{L}{\cdot}\mathbf{M}]_{\text{amide}}$, and, thus, results in a lower apparent K_{a} for the pseudorotaxane formation.

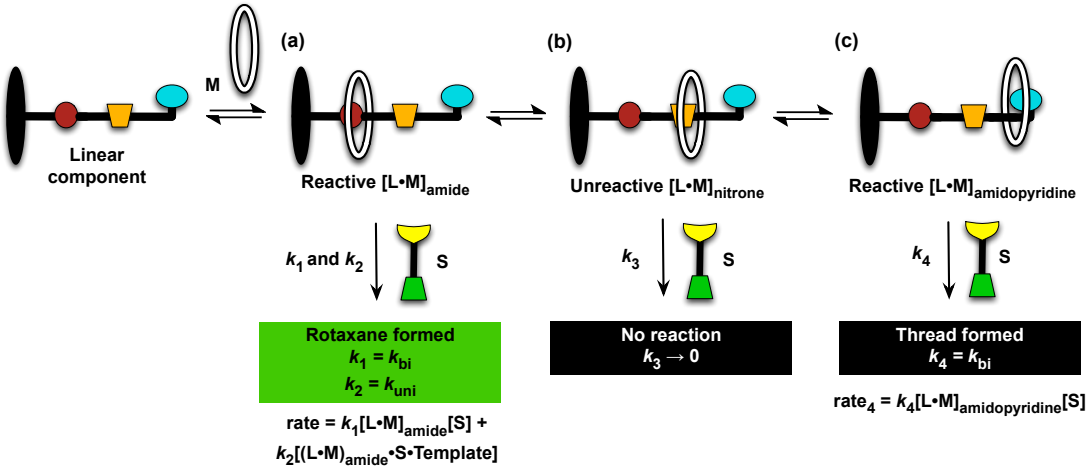


Figure 6.27 Linear component **L** can associate with a macrocycle **GM** and **PM** (represented as **M**) to give **(a)** a reactive pseudorotaxane complex with the macrocycle bound at the desired amide binding site. Alternatively, the macrocycle can also adopt two undesirable co-conformations with the linear component, forming an **(b)** unreactive pseudorotaxane with **M** bound at the nitrone reactive site and **(c)** reactive pseudorotaxane with **M** at the amidopyridine site, which can react with stopper **S** to form thread, through a slow bimolecular reaction. Figure adapted with permission from Ref. 301. Copyright 2015 American Chemical Society.

The apparent K_{a} for the pseudorotaxane formation reflects[e] the binding of the macrocycle to three different binding sites. Similarly, the observed unimolecular rate constant (k_{uni}) associated with the template-mediated formation of each rotaxane is lower than it would be if the macrocycle was bound solely at the amide binding site.

The rotaxane–thread replicating systems are highly interconnected, with numerous catalytic as well as non-catalytic complexes present in each reaction mixture simulta-

[e]Throughout the fitting of rotaxane kinetic data, only a single K_{a} parameters was fitted for the formation of all three pseudorotaxane complexes: $[\mathbf{L}{\cdot}\mathbf{M}]_{\text{amide}}$, $[\mathbf{L}{\cdot}\mathbf{M}]_{\text{nitrone}}$ and $[\mathbf{L}{\cdot}\mathbf{M}]_{\text{amidopyridine}}$. It is possible, in fact likely, that the K_{a} for the formation of the three pseudorotaxane might vary slightly from each other, however, it was not possible to deconvolute the three individual binding events. Therefore, in order to avoid fitting a larger number of variables than data sets available, only a single K_{a} value was employed in the fitting process for each rotaxane system.

neously. In order to develop a better understanding of how the system is affected by the presence of the additional macrocycle binding sites, *i.e.* the nitrone unit and the amidopyridine unit, fitting of the concentration *vs* time data available from the kinetic experiments was undertaken. The kinetic data from the related thread and rotaxane kinetic experiments were fitted to kinetic models encompassing all interactions and reactions for each system. The rotaxane models included also the formation of the three different pseudorotaxane complexes ($[\mathbf{L}\cdot\mathbf{M}]_{\text{amide}}$, $[\mathbf{L}\cdot\mathbf{M}]_{\text{nitrone}}$ and $[\mathbf{L}\cdot\mathbf{M}]_{\text{amidopyridine}}$). Utilising an appropriate fitting procedure (for an example script, see **Appendix A5**), it was possible to extract normally inaccessible kinetic parameters, namely, the template-mediated unimolecular rate constants (k_{uni}) and the product duplex association constants ($K_{\text{a}}^{\text{Duplex}}$) for the formation of **T**, **GM-R** and **PM-R** (**Table 6.1**). In order to avoid fitting a larger number of constants than the data sets available, two assumptions were implemented throughout the fitting procedure: (i) the bimolecular rate constant and the $K_{\text{a}}^{\text{Duplex}}$ were assumed to be identical for the formation of all three templates and (ii) only a single unimolecular rate constant was fitted for each template across the auto- and crosscatalytic kinetic experiments, leading to the formation of a particular template. The latter assumption was implemented based on the closely matched catalytic efficiencies of the auto- and crosscatalytic pathways determined during the kinetic experiments. The assumption was further supported by the fact that attempts to fit different unimolecular rate constants for the individual auto- and crosscatalytic pathways revealed little variation in the fitted values. Interestingly, the fitting revealed that k_{uni} for the formation of both rotaxanes are, as a result of the various binding sites present on the linear component (only one leading to formation of rotaxane) almost four times smaller that the k_{uni} for the formation of the thread. The decrease in the k_{uni} in the rotaxanes reflects the decrease in the reactivity of the rotaxane-forming, catalytically-active quaternary complexes relative to the thread-forming ternary complexes as a result of the multiple binding sites. Using the fitted values of k_{bi} and k_{uni}, the kinetic effective molarity ($\text{EM}_{\text{kinetic}}$) for the formation of each template was calculated. This parameter provides a measure of the efficiencies of the template-mediated catalytic pathways leading to the formation of thread and both rotaxanes, relative to the bimolecular pathways ($\text{EM}_{\text{kinetic}} = k_{\text{uni}}/k_{\text{bi}}$). The apparent K_{a} values determined for the formation of [**L**·**GM**] and [**L**·**PM**] pseudorotaxane complexes (**Table 6.1**) reflect the trend in association constants previously determined through low temperature experiments for [**N**·**M**].

Analysis of the fitted kinetic parameters for thread and rotaxanes (**Table 6.1**) shows the detrimental effect of the additional of undesirable macrocycle-binding sites on **L** on the apparent K_{a} for the formation of reactive pseudorotaxane complex and the unimolecular rate constant and therefore, also the ratio of [**R**]/[**T**] formed in the system.

Table 6.1 Kinetic parameters and association constants obtained using SimFit for the formation of **T**, **GM-R** and **PM-R** through kinetic fitting of the experimental kinetic data (at 5 °C). In all cases, the k_{bi} and K_a^{Duplex} were determined as 1.04×10^{-4} M^{-1}s^{-1} and 10.3×10^{6} M^{-1} for **T**, **GM-R** and **PM-R**. Table taken with permission from Ref. 301. Copyright 2015 American Chemical Society.

	Thread	GM-R	PM-R
k_{uni} / 10^{-4} s^{-1}	226	61.0	70.4
$EM_{kinetic}$ / M	217	59	68
K_a [L•M] / M^{-1}	–	220	344

In order to demonstrate how a similar system, integrating the formation of a rotaxane with self-replication, would work in the absence of competitive binding sites, a series of kinetic simulations was performed. These simulations (see **Appendix A5** for an example script) employed the parameters determined for **PM-R** system, which showed more efficient rotaxane formation than **GM-R** as a result of a higher $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$. The simulations were designed specifically to explore the effect of varying the strength of [**L**·**M**] association constant from 10^2 to 10^5 M^{-1} on the ratio of [**R**]/[**T**] under two different conditions: condition A employed kinetic parameters determined for the **PM-R** experimental system (k_{uni} rotaxane $<$ k_{uni} thread); condition B examined an ideal version of the **PM-R** system, where **L** possessed only the desired macrocycle binding site (k_{uni} rotaxane = k_{uni} thread). Simulating the condition B (**Figure 6.28**), where the effective molarity for the thread and the rotaxane are identical, enabled elucidation of the behaviour of a rotaxane–thread system in a situation where rotaxane formation would not be hindered by competitive binding sites—a situation where the ratio of products formed depends solely on the K_a governing the formation of [**L**·**M**].

The results of the simulation in condition A (**Figure 6.28**) revealed that an increase in $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ does not afford a ratio of [**R**]/[**T**] greater than 8 in the condition employing parameters determined for **PM-R** system, unless $K_a^{[\mathbf{L}\cdot\mathbf{M}]} > 10^5$ M^{-1}. While a considerable improvement over the current product ratios, it is possible that a larger K_a value for the formation of pseudorotaxane [**L**·**M**] might also result in a concurrent increase in the strength of association for the macrocycle binding sites (see **Figure 6.27**), therefore, resulting in a lower [**R**]/[**T**] ratio than predicted from this simulation. Whilst the current experimental design, incorporating two additional competing macrocycle-binding sites, does not favour rotaxane formation, the same kinetic model can afford significantly higher [**R**]/[**T**] ratios (higher than 30) if only a single macrocycle binding site is present (**Figure 6.28**, condition B), demonstrating that rotaxane can be formed selectively in a replicating network exploiting orthogonal recognition processes (based on Model 2), and, thus, operating simultaneously through both the auto- and the crosscatalytic pathways.

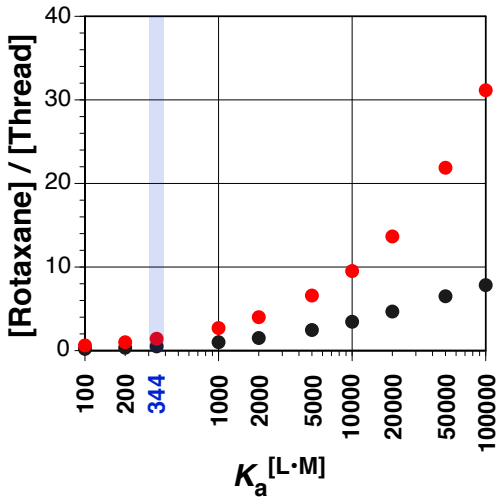


Figure 6.28 Outcome of kinetic simulations looking at how increasing the $K_a^{[L\cdot M]}$ affects the ratio of [**R**]/[**T**] formed in the replicating network under two conditions: condition A (black circles) employing the kinetic parameters obtained from fitting of kinetic experimental data (k_{uni} rotaxane $< k_{uni}$ thread) and condition B (red circles) examining an idealised situation where the system does not have additional undesired macrocycle-binding sites on the linear component (k_{uni} rotaxane = k_{uni} thread). The $K_a^{[L\cdot M]}$ value of 344 M^{-1} for [**L**·**PM**] (blue rectangle) represents the outcome determined for the condition employing kinetic parameters for **T** and **PM-R**. Simulations were performed using SimFit software package. Figure adapted with permission from Ref. 301. Copyright 2015 American Chemical Society.

6.5 Model 3

The efficiency of the replicating system, successfully integrating rotaxane formation with self-replication processes based on Model 2, suffered as a result of several issues. Firstly, the orthogonal recognition processes employed resulted in self-replication taking place at the periphery of the structure, and, thus, the presence or absence of a macrocycle made little difference to the reaction itself. The location of the macrocycle within the molecular framework allowed formation of a replicating network where the formed templates, thread and rotaxane, exhibited the same capacity for auto- and crosscatalysis. As a result of the crosscatalytic relationship between the thread and the rotaxane, the ratio of these species depends on the K_a for the association between the linear component and macrocycle. The second factor influencing this particular design is the presence of additional macrocycle binding sites that permit formation of undesirable pseudorotaxane complexes between linear component and macrocycle (**Figure 6.27**). Kinetic simulations illustrated that in a system where these undesirable binding sites are absent, the rotaxane can outperform the thread significantly ([**R**]/[**T**] ratio > 30 at high values of $K_a^{[L\cdot M]}$).

A replicating network with the capacity to form a [2]rotaxane preferentially can be also designed using a different kinetic framework. In particular, Model 3 (**Figure 6.29a**),

exploiting a macrocycle equipped with one of the recognition sites required for self-replication presents a suitable design where rotaxane formation should be favoured. Specifically, rotaxane formation is likely to be significantly more efficient than that of the thread, as the rotaxane is the only component produced in the network with the capacity to take part in any template-mediated catalytic processes.

In order to probe the potential of Model 3 for the development of a network capable of forming a [2]rotaxane selectively, the recognition and reaction processes within this model (**Figure 6.29a**) were translated into a kinetic simulation script (see **Appendix A5** for an example). The kinetic and recognition parameters employed in this model, namely the k_{bi} (**Figure 6.29a**) and ${K_a}^{Ind}$ (**Figure 6.29b**), were based on the parameters determined through kinetic fitting for the PM [2]rotaxane. The rate constant for the recognition-mediated reaction (**Figure 6.29a**, k_{uni}) was set to a value that produces an $EM_{kinetic}$ (**Figure 6.29c**) of 10 M—a reasonable value for catalytic efficiency that is at the lower end of the spectrum of catalytic efficiencies determined for the range of self-replicating systems designed by the Philp laboratory to date. The strength of the duplex association is harder to predict, and, therefore, three different conditions (I, II and III) of EM_{thermo} (**Figure 6.29c**) were screened: 0.1, 1.0 and 10 M. At the same time, the K_a for the formation of the pseudorotaxane complex was varied (${K_a}^{[\mathbf{L}\cdot\mathbf{M}]}$ = 100, 200, 500, 1000 and 2000 M^{-1}), with the view to examining how these parameters affect the ratio of [rotaxane]/[thread] formed within the system, as well as the overall % conversion to all cycloadducts within each simulated condition (relative to $[C]_{initial}$ = 10 mM).

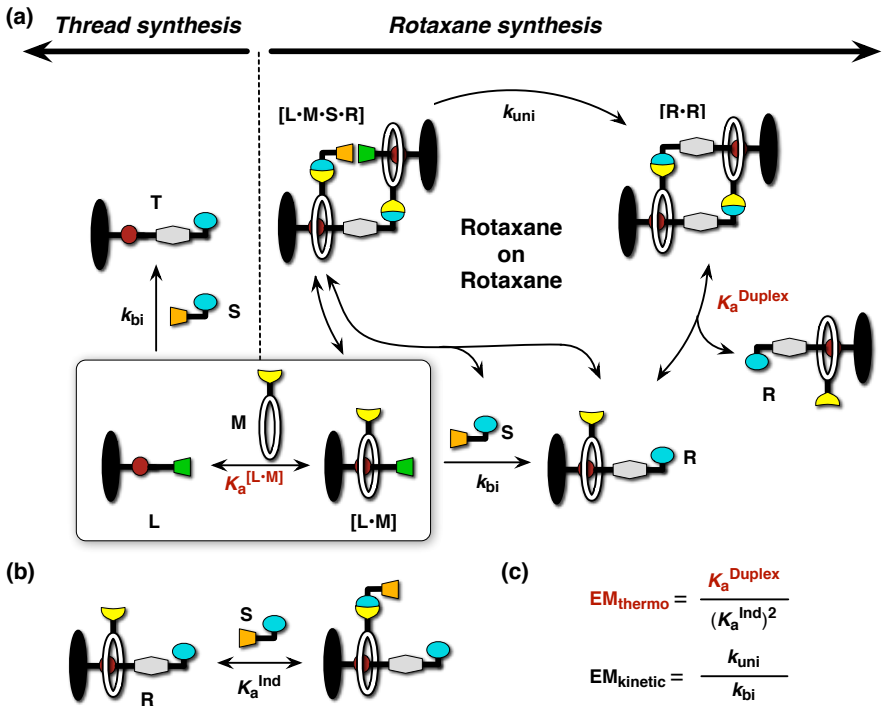


Figure 6.29 **(a)** Cartoon representation of Model 3, where macrocycle **M** bears one of the recognition sites required for self-replication. The replicating system is based around a key equilibrium (solid rectangle, bottom left) between a linear component **L** and macrocycle **M**, to produce a pseudorotaxane complex [**L**·**M**]. From the central equilibrium, the linear component can react with stopper **S** only *via* template-independent bimolecular pathway to produce thread **T**. Pseudorotaxane [**L**·**M**] can react with **S** through both the bimolecular reaction and through recognition-mediated self-replicating pathway driven by catalytically-active complex [**L**·**M**·**S**·**R**] to form [2]rotaxane **R**. For simplicity, equilibrium arrows are shown as double-headed arrows. The yellow and blue cartoons represent the recognition sites and green and blue the reactive sites needed for self-replication. Grey shape represents the reaction product and the macrocycle-binding site is shown in red. **(b)** Example of single point association between an amidopyridine and a carboxylic acid, governed by K_a^{Ind}. **(c)** Equations representing $EM_{kinetic}$ and EM_{thermo}. Rate and association constants varied throughout the simulations are highlighted in red.

Analysis of the simulation outcome of the EM_{thermo} conditions revealed several interesting trends in both the [**R**]/[**T**] ratio (**Figure 6.30a**) as well as the % conversion[f] to products (**Figure 6.30b**). Looking at the ratio of [**R**]/[**T**] across the three conditions of EM_{thermo} showed that the rotaxane is formed with the highest efficiency when the EM_{thermo} is the lowest, *i.e.* 0.1 M. Even at an early reaction time (four hours) and a low value of $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ (100 M^{-1}), the [**R**]/[**T**] ratio remained above 4. The ratio increased dramatically at $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ value of 2000 M^{-1}, reaching a value $>$ 160 after four hours and $>$ 350 after eight hours (**Figure 6.30Ia**). In all three conditions of EM_{thermo}

[f]The % conversion to all products as opposed to the % of the rotaxane within the product pool was examined as the rotaxane is the dominant product in each case—even in the least efficient scenario examined, condition III where the EM_{thermo} and EM_{thermo} = 10, the rotaxane accounted for over 50% of the product pool.

(**Figure 6.30a**), the ratio of [**R**]/[**T**] was found to be higher than the ratio determined in the kinetic simulations examining the **PM-R** rotaxane system. In fact, in no instance did the [**R**]/[**T**] ratio fall below 1.5. The increase in EM_{thermo} from 0.1 to 1.0 and 10 M in conditions II and III resulted in an overall decrease in the [**R**]/[**T**] ratio at each $K_a^{[L \cdot M]}$ condition examined. This decrease in the efficiency of rotaxane formation results from the higher template duplex stability, which in turn has a negative impact on the catalytic turnover. Nevertheless, in both condition II and III, the [**R**]/[**T**] ratio after eight hours was higher than after four hours, suggesting that, while the increase in the [**R**]/[**T**] ratio is slower than in condition I as a result of product inhibition, the system continues to make the rotaxane product preferentially.

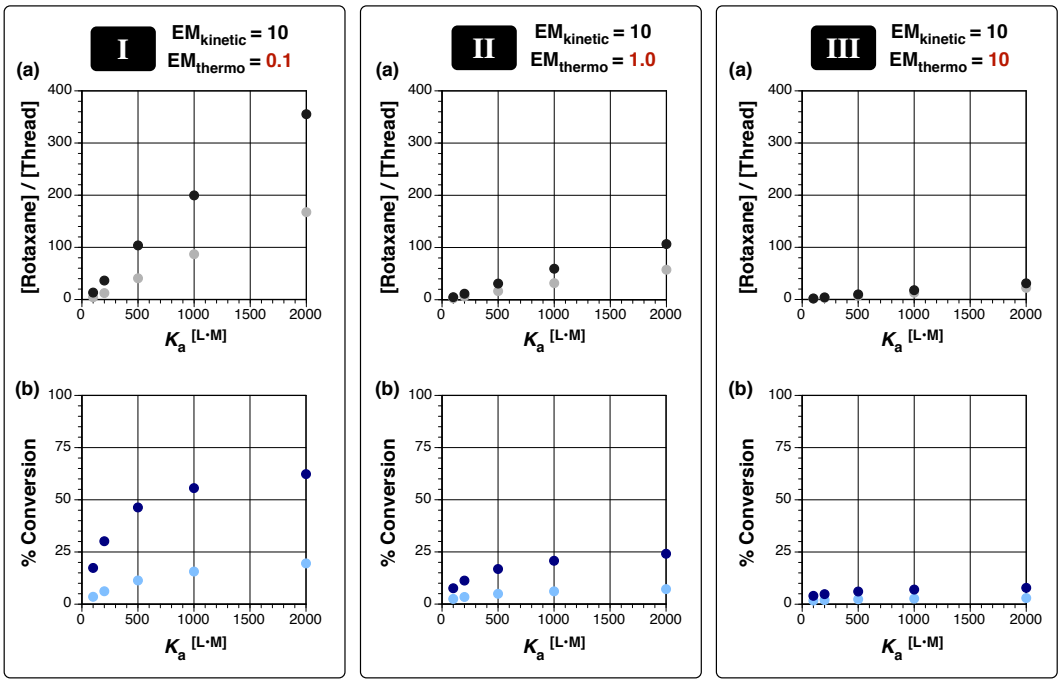


Figure 6.30 Outcome of kinetic simulations looking at how increasing $K_a^{[L \cdot M]}$ affects **(a)** the ratio of [**R**]/[**T**] and **(b)** % conversion to all cycloadducts formed in a self-replicating rotaxane network based on Model 3, after four hours (black/dark blue circles) and after eight hours (grey/light blue circles). In each simulation, the initial concentration of reactants is set to 10 mM. The simulations examined three conditions of EM_{thermo}, I to III: 0.1, 1.0 and 10 M, whilst keeping the $EM_{kinetic}$ (10 M) and the $k_{bimolecular}$ identical to that determined through kinetic fitting for **PM-R**.

Examination of the % conversion in the three simulated EM_{thermo} conditions showed a trend similar to that observed for the [**R**]/[**T**] ratio. The combined formation to all products proceeds with highest efficiency in condition I and decreases progressively in II and III. In condition I (EM_{thermo} = 0.1 M), conversion after eight hours reached $> 60\%$. In comparison, the conversion after the same time in condition III reached a mere 8%. The nature of the Model 3 is such that despite the low overall conversion, the rotaxane is the only product that can be formed *via* the template-driven pathway. Therefore,

albeit the slow conversion, the rotaxane product remains the dominant product, and its proportion in the product mixture continues to increase over time.

The three simulated systems operating at EM_{thermo} from 0.1 to 10 M probed the effect of template duplex stability on the efficiency of rotaxane formation within the replicating thread–rotaxane network. Only the rotaxane product is capable of self-replication, and, thus, increase in the $EM_{kinetic}$ to a higher value would simply increase the efficiency of rotaxane formation, resulting in a higher overall % conversion to products as well as a higher ratio of [**R**]/[**T**] in each condition—and the simulations were therefore not undertaken. In the next step, one of the simulated conditions was selected as a platform for examination of the effect of initial reaction concentration ($[C]_{Initial}$) on the % conversion to all products, as well as the ratio of of [**R**]/[**T**]. Specifically, this simulation examined a system based on a linear component, macrocycle and stopper at various concentrations ($[C]_{Initial}$ = 0.1, 0.5, 2.5, 5, 10, 25 and 50 mM) with an association constant for the formation of the pseudorotaxane set at 200 M^{-1}. The EM_{thermo} and $EM_{kinetic}$ for this series of simulations remained fixed throughout at 0.1 and 10 M, respectively (**Figure 6.31**).

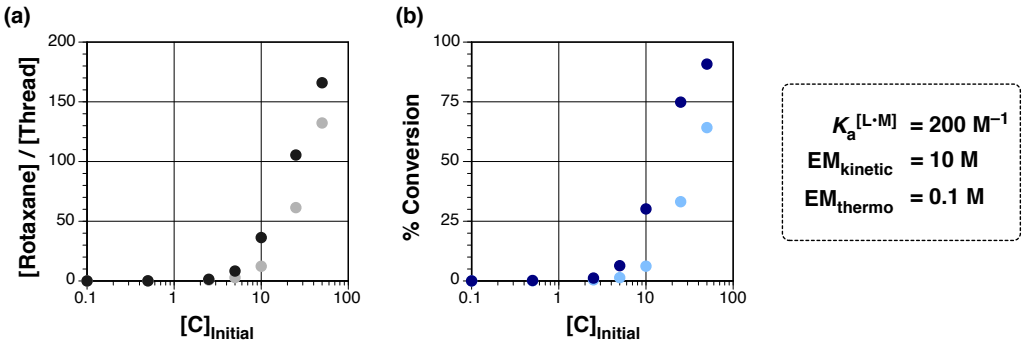


Figure 6.31 **(a)** The ratio of [**R**]/[**T**] and **(b)** % conversion to all cycloadducts formed in a simulated self-replicating rotaxane network (Model 3), after four hours (black/dark blue circles) and after eight hours (grey/light blue circles) at initial reaction concentrations = 0.1, 0.5, 2.5, 5, 10, 25 and 50 mM.

The outcome of these kinetic simulations examining the effect of a range of reaction concentrations on the efficiency of rotaxane formation showed that below 5 mM, both the [**R**]/[**T**] ratio (**Figure 6.31a**) and % conversion remain very low at both four and eight hours (**Figure 6.31b**). At concentrations above 5 mM, rotaxane efficiency increases fairly rapidly. The K_a^{Ind} in the system is 910 M^{-1}, which means that the K_d for the individual recognition event is just above 1 mM. The marked decrease in the formation of the rotaxane at concentrations < 5 mM in both sets of simulations stems from the decreased ability of the recognition processes to operate efficiently. While the simulation varied the concentration at which the system performed, the K_a for the formation of pseudorotaxane complex [**L**·**M**] was maintained at 200 M^{-1}. The

simulation outcome showed that the [**R**]/[**T**] ratio can be increased not only by an increase in $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$, but also by an increase in the concentration at which the reaction is undertaken. Higher reaction concentrations result in a higher proportion of the desired reactive pseudorotaxane in the mixture, which in turn allows the [2]rotaxane to form more efficiently. The outcome of these kinetic simulations is clear: the design of a reaction network based on Model 3 has the potential to produce a system where rotaxane formation is favoured significantly ([**R**]/[**T**] ratio of almost 400). In particular, the kinetic simulations revealed that the design will operate the most efficiently at high reaction concentrations and under conditions where $\mathrm{EM}_{\mathrm{kinetic}}$ and $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ are as high as possible and the $\mathrm{EM}_{\mathrm{thermo}}$ is as low as possible.

The level of selectivity for the rotaxane product over thread determined in in these kinetic simulations based on Model 3 (**Figure 6.32b**) is in stark contrast to the [**R**]/[**T**] ratios obtained in simulations of a network constructed using Model 2 (**Figure 6.32a**). Examination of the [**R**]/[**T**] product ratios achieved at four different values of $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ (100, 200, 1000 and 2000 M^{-1}) in kinetic simulations examining Model 2 revealed that even in the absence of competitive macrocycle binding sites, the maximum ratio of [**R**]/[**T**] does not exceed 4 after eight hours. The ratios in the simulations employing Model 3, on the other hand, reach a ratio as high as 350 in the condition employing the lowest $\mathrm{EM}_{\mathrm{thermo}} = 0.1$ M and never fall below 2 (even in the condition employing the highest $\mathrm{EM}_{\mathrm{thermo}} = 10$ M), after eight hours.

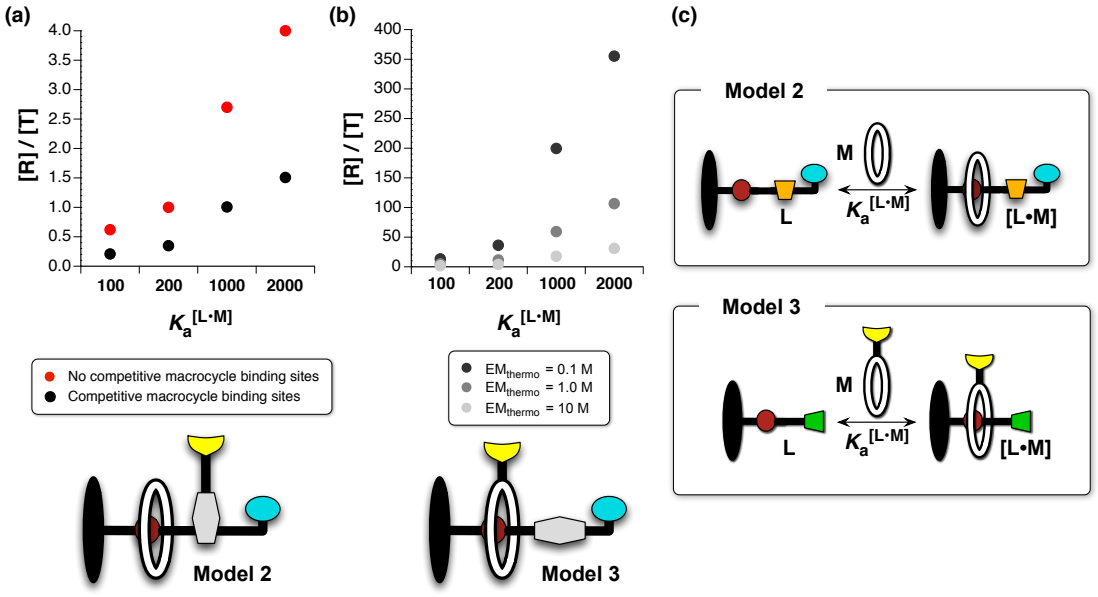


Figure 6.32 Comparison of the [**R**]/[**T**] product ratios obtained in kinetic simulations examining the effect of varying $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ in a reaction–thread network based on **(a)** Model 2 and **(b)** Model 3, determined after eight hours. Initial reaction concentration was set to 10 mM. **(c)** Comparison of the central equilibria between a linear component **L** and pseudorotaxane complex [**L**·**M**] in Model 2 and Model 3.

Clearly, Model 3 possesses a significant advantage over Model 2 in its ability to form the rotaxane component preferentially. In Model 2, the central equilibrium (**Figure 6.32c**) is formed by the free linear component **L** and the pseudorotaxane complex [**L**·**M**], both of which can participate in the respective thread and rotaxane autocatalytic cycles, leading to the non-linear amplification of both products. In contrast, only the pseudorotaxane complex is capable of taking part in template-directed reactions—this means that the pseudorotaxane formed within the system will be removed from the central equilibrium much faster than the linear component, leading to significant amplification of the rotaxane product.

In order for the experimental implementation of a reaction network based on Model 3 to achieve the potential demonstrated through these kinetic simulations, design of the components comprising the system will need to avoid the shortcomings of the previous models. Namely, the design should exploit building blocks equipped with sufficiently large stoppering groups that will prevent formation of any undesirable pseudorotaxane complexes—a process that can reduce the efficiency of rotaxane formation significantly. The macrocycle has the capacity to move along the threading component within a rotaxane, and, therefore, the linear component integrating the macrocycle-binding site should, in particular, avoid incorporation of the reactive nitrone site, capable of strong hydrogen-bonding interactions. The molecules required for the formation of the rotaxane should also possess sufficient structural rigidity, in order to promote the template-directed pathway, relative to the binary complex route. A preliminary design that is close to fulfilling the above-described requirements, for the experimental examination of a rotaxane-forming network based on Model 3, constructed from a linear component **167**, macrocycle **168** and stopper **169**, is portrayed in **Figure 6.33**.

Fortuitously, the linear component **167** for use in this network based on Model 3 can be prepared (**Scheme 6.13**) from an already available building block, **N**, that was used previously in the synthesis of linear component **L**. This —NO_2 group bearing intermediate **N** is already equipped with the desired macrocycle binding site. Reduction of **N** using Pd/C (10% wt) and sodium hypophosphite in aqueous THF produced the amine **170** in 89% yield. Reaction of amine **170** with maleic anhydride produced the intermediate compound **171** which was cyclised immediately in acetic acid under microwave heating, producing the final maleimide **167** in a 29% yield over the two steps.

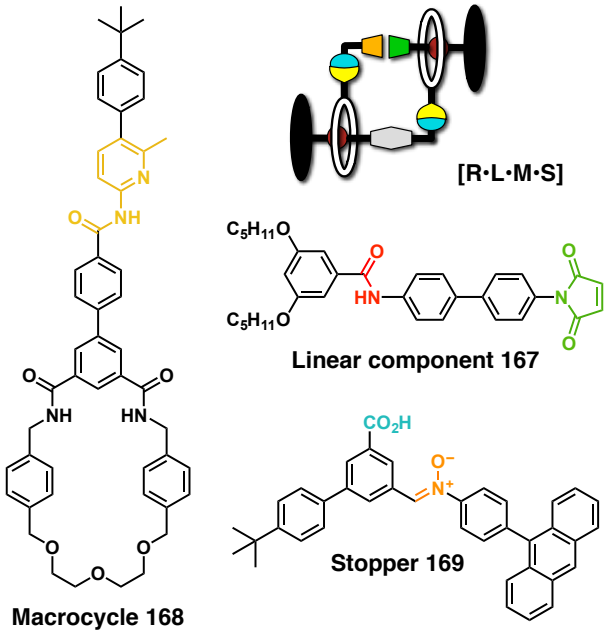
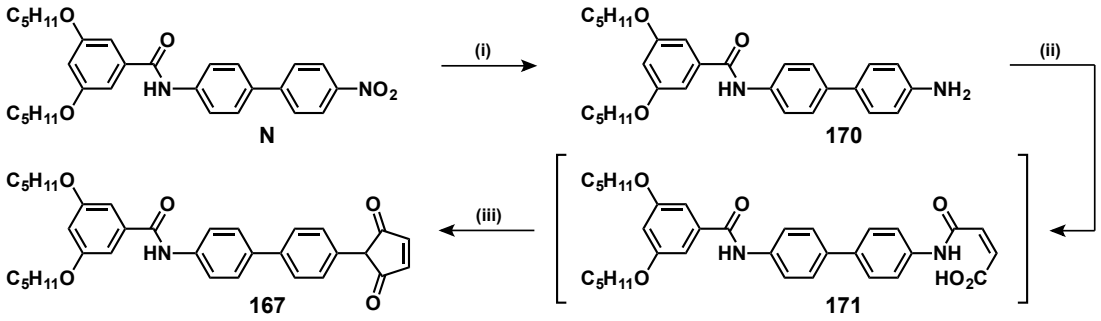


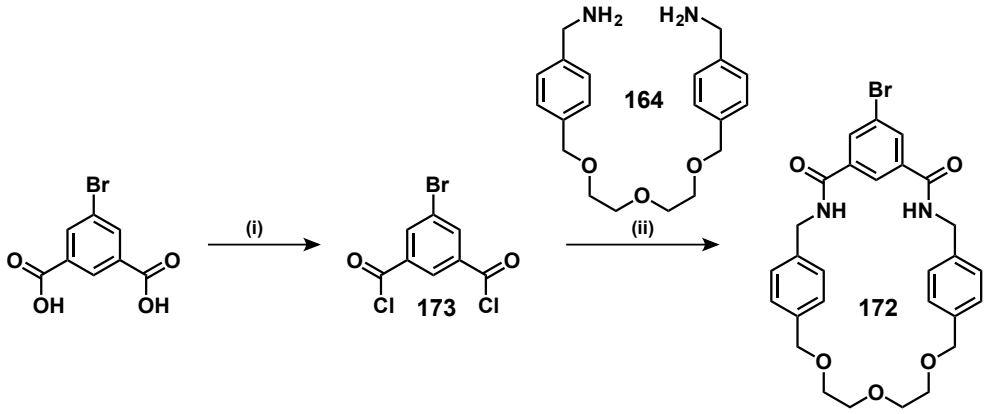
Figure 6.33 Cartoon and chemical structure representation of an interconnected reaction network integrating replication processes with [2]rotaxane formation based on Model 3. The network is assembled form a linear component **167**, a macrocycle **168** and a stopper **169**. These three components bear recognition sites (blue and yellow) and reactive sites (orange and green) required for self-replication. Macrocycle-binding site is shown in red. The building blocks are also equipped with bulky *t*-butylbenzene or anthracene groups in order to prevent formation of undesirable pseudorotaxane complexes.



Scheme 6.13 Conditions and reagents employed in the synthesis of maleimide **167**. (i) Sodium hypophosphite, Pd/C (10% wt), THF/H_2O, 40 °C, 5 h, 89%. (ii) maleic anhydride, THF, quantitative conversion, (iii) AcOH, 120 °C in MW, 2 h 29%.

In contrast to macrocycles **GM** and **PM** employed thus far, the new macrocycle must incorporate the recognition group that will permit the formation of a self-replicating rotaxane. To this end, the pyridine moiety present in the **GM** macrocycle was substituted[304] with a bromophenyl ring that permits introduction of the recognition-appendage through Suzuki or Sonogashira coupling. Preparation of the key, intermediate bromomacrocycle **172** started with the reaction of 5-bromoisophthalic acid in thionyl chloride to produce the activated 5-bromoisophthaloyl dichloride **173**. Reaction of the activated

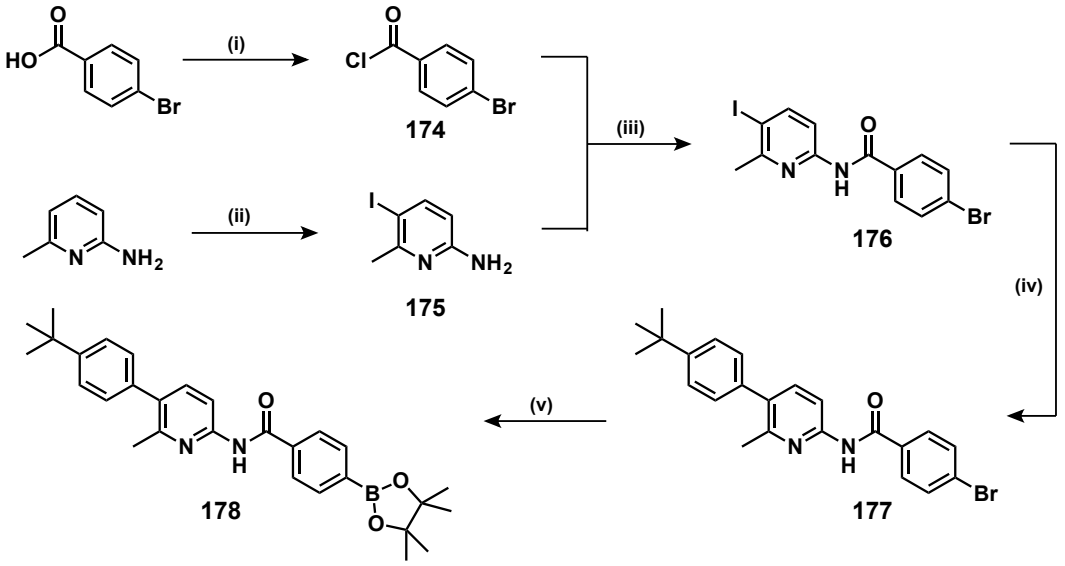
intermediate **173** with a previously employed glycol diamine **164** under high-dilution conditions produced the target macrocycle **172** in a good 20% yield (**Scheme 6.14**).



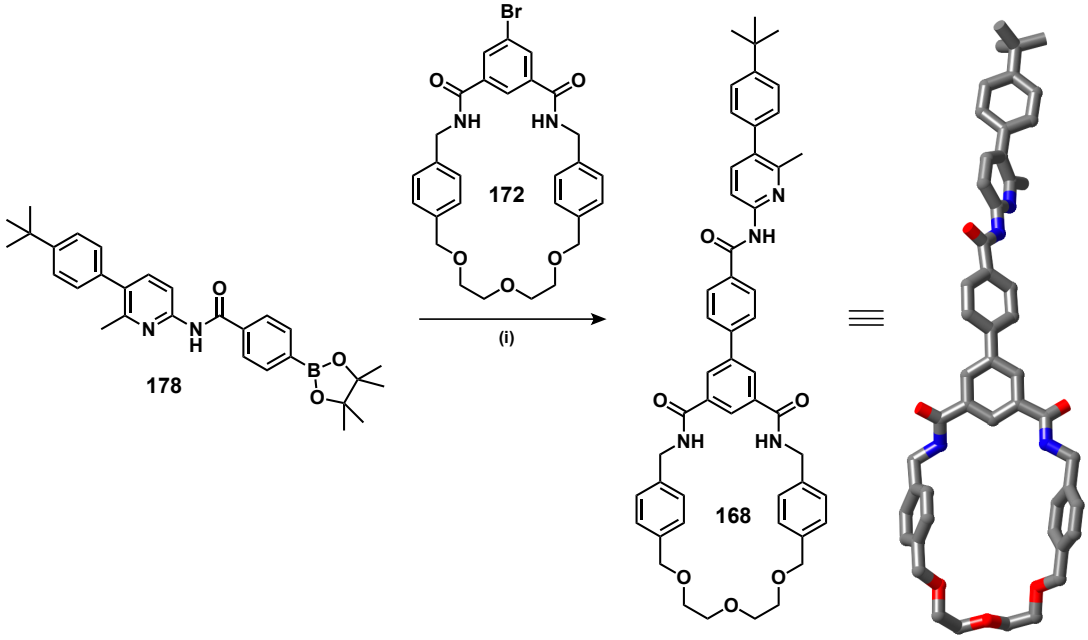
Scheme 6.14 Conditions and reagents employed in the synthesis of macrocycle precursor **172**. (i) $SOCl_2$, PhMe, 100 °C, 6 h, quantitative conversion. (ii) CH_2Cl_2, Et_3N, RT, five days, 20%.

The next step in the synthetic route was the preparation of the macrocycle recognition arm (**Scheme 6.15**). In the first step, a 4-bromobenzoic acid was treated with thionyl chloride to give the activated analogue **174**. The crude acyl chloride was reacted with an iodinated aminopyridine **175**, prepared by iodination of 2-amino-6-methylpyridine. The furnished compound **176**, bearing the 6-amino amidopyridine recognition site, was selectively coupled with 4-*tert*-butylbenzeneboronic acid to give the compound **177**, now equipped with a bulky *tert*-butylphenyl group stopper. Using bis(pinacolato) diboron, KOAc and a palladium catalyst under microwave irradiation, the bromo compound **177** was converted to a pinacol ester **178**, which was used directly in the next step without purification.

The final macrocycle **168** (**Scheme 6.16**) was prepared through Suzuki coupling of the two key intermediate compounds, **172** and **178**, using $CsCO_3$ as a base, in a 38% yield. Single crystals of the *tert*-butylbenzene-amidopyridine macrocycle **168**, suitable for analysis by X-ray diffraction, were obtained by slow evaporation from a solution of CH_2Cl_2/MeCN. In the past, computational modelling of recognition-enabled macrocycles showed[304,314–316] that substitution of the pyridine unit within the **GM** macrocycle with an isophthalic acid building block increases the size of the macrocycle cavity significantly, however, the structure of **168** in the solid state (**Scheme 6.16**), whilst noticeably more disordered than the pseudo-chair conformation adopted generally by **GM**, did not reveal a marked increase in the volume of the cavity.



Scheme 6.15 Conditions and reagents employed in the synthesis of macrocycle precursor **178**. (i) $SOCl_2$, PhMe, 150 °C, 3 h, quantitative conversion. (ii) I_2, HIO_4, $AcOH/H_2SO_4$, 60 °C, 20 h, 43%. (iii) CH_2Cl_2, Et_3N, 0 °C to RT, 39%. (i) 4-*tert*-butylbenzeneboronic acid, Na_2CO_3, $Pd(PPh_3)_4$, $EtOH/PhMe/H_2O$ (1 : 2 : 1), 50 °C, 2 days, 71%. (v) bis(pinacolato)diboron, $PdCl_2(dppf)\cdot CH_2Cl_2$, KOAc, dioxane, 120 °C (MW), 3 h, quantitative conversion.



Scheme 6.16 Conditions and reagents employed in the synthesis of macrocycle **168**: (i) $CsCO_3$, $EtOH/PhMe/H_2O$ (3:2:2), 90 °C, 2 days, 38%, and a stick representation of the X-ray crystal structure obtained for macrocycle **168** by X-ray crystallographic analysis of single crystals. C atoms are coloured grey, O atoms in red and N atoms in blue. Hydrogen atoms are omitted for clarity. Data collection and X-ray diffraction analysis were performed by Dr. D. B. Cordes and Professor A. M. Z. Slawin at the University of St Andrews.

Examination of macrocycle **168** by ^{1}H NMR spectroscopy revealed an interesting behaviour—at low concentrations, the individual signals corresponding to the macrocycle protons could be clearly identified, however, at higher concentrations, the ^{1}H NMR spectrum of **168** becomes significantly more complex, suggesting presence of both the free macrocycle and the product of its association to form daisy chain-like[317–319] assemblies, where the amidopyridine unit in one macrocycle **168** is threaded through the cavity of another. The ability of this macrocycle to self-associate at higher concentrations indicates that the isophthalic acid does indeed increase the cavity size—the cavities of **GM** and **PM** macrocycles examined thus far have been established through binding studies to be too small to permit association with compounds bearing a bulky group such as the 2-methyl-3-*tert*-butylbenzene amidopyridine present in macrocycle **168**. A preliminary binding study, examining an equimolar (5 mM) solution of macrocycle **168** and maleimide linear component **167** after 24 hours by 499.9 MHz ^{1}H NMR ($CDCl_3$, 22 °C) revealed that the exchange between the bound and the unbound state is in a slow exchange regime on the NMR chemical shift time-scale. Using the resonances corresponding to the linear maleimide in the bound and free state, it was possible to determine the K_a for this interaction as 1670 M^{-1} using a single-point method. The high association constant for the formation of pseudorotaxane [**167**·**168**] is highly promising, and will hopefully lead to construction of an efficient and highly selective [2]rotaxane based on Model 3 in the future.

6.6 Summary

This chapter presented the design, successful experimental implementation and full characterisation of an interconnected reaction network integrating self-replication processes with a second, recognition-directed process, leading to the formation of a mechanically-interlocked architecture—a [2]rotaxane. Exploiting orthogonal recognition processes, the reaction network, assembled from a linear component, macrocycle and a stopper, was effectively directed to form a thread (**T**), lacking the macrocyclic component, and a [2]rotaxane (**R**). Comprehensive kinetic analysis of the reaction pathways available to the network revealed that the thread and rotaxane products are matched in catalytic efficiencies as templates, in both auto- and crosscatalytic cycles (**Figure 6.34a**).

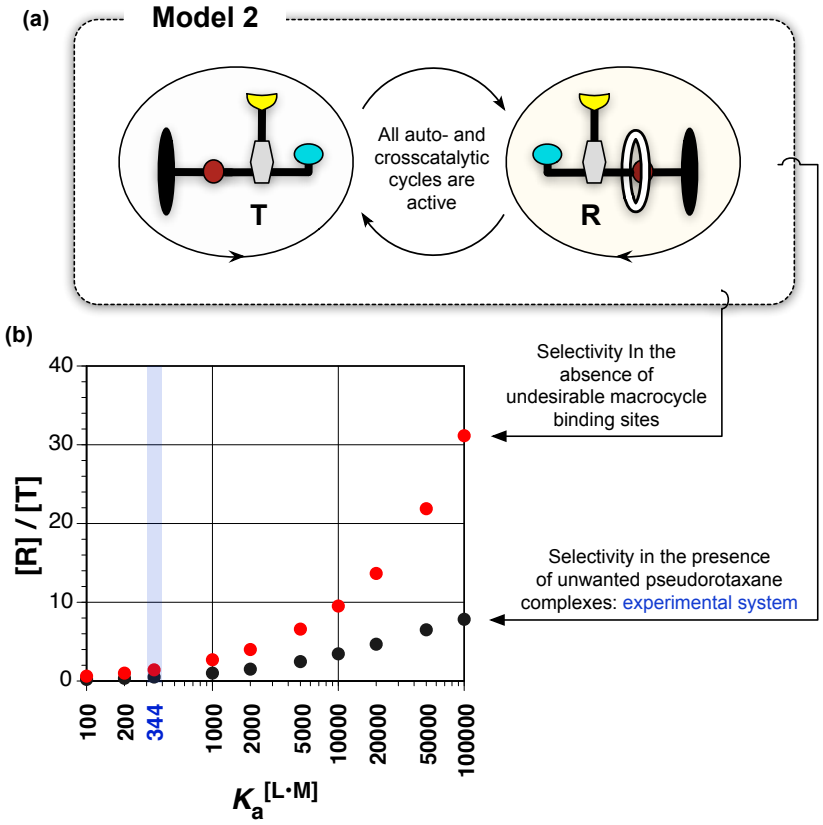


Figure 6.34 **(a)** Orthogonal recognition processes allowed construction of a rotaxane–thread network based on Model 2, where all auto- and crosscatalytic pathways are active—rotaxane **R** and thread **T** template the formation of each other with efficiency equal to that in the autocatalytic pathways. **(b)** The design of the employed rotaxane forming components permits formation of several undesirable pseudorotaxane complexes which allow the formation of thread to proceed more efficiently than formation of the rotaxane. Kinetic simulations showed than in the absence of these additional binding sites capable of associating with the macrocycle, the ratio of [**R**]/[**T**] in the reaction network can increase significantly relative to the situation where formation of unproductive pseudorotaxane complexes is possible.

However, as a result of the interplay between the various recognition and reaction processes operating in the network simultaneously, specifically the nitrone and amidopyridine moieties acting as competitive binding sites for the macrocyclic component, the thread formation surpassed rotaxane formation in the experimental system based on Model 2. Employing kinetic simulations, it was possible to demonstrate that in the absence of the competitive macrocycle binding sites, which permit formation of undesirable pseudorotaxane complexes, orthogonal recognition processes can be successfully exploited to drive the three-component reaction network to make the [2]rotaxane product preferentially.

Model 3, owing to its unique design strategy where the macrocycle bears one of the recognition sites required for self-replication, presents a reaction network that can produce [2]rotaxane, as the only product capable of taking part in template-mediated

reactions, selectively, from a mixture of components. The reactivity in such a rotaxane–thread network composed of a linear component, a macrocycle and a stopper, and in particular the possibility of producing a [2]rotaxane preferentially, was probed through a series of kinetic simulations examining how the strength of the rotaxane duplex and pseudorotaxane complex, and reaction concentration. The simulation outcome showed that at optimised reaction conditions, the reaction network forms the [2]rotaxane interlocked architecture with excellent selectivity, affording a high ratio of [**R**]/[**T**] (as high as 350, **Figure 6.35**).

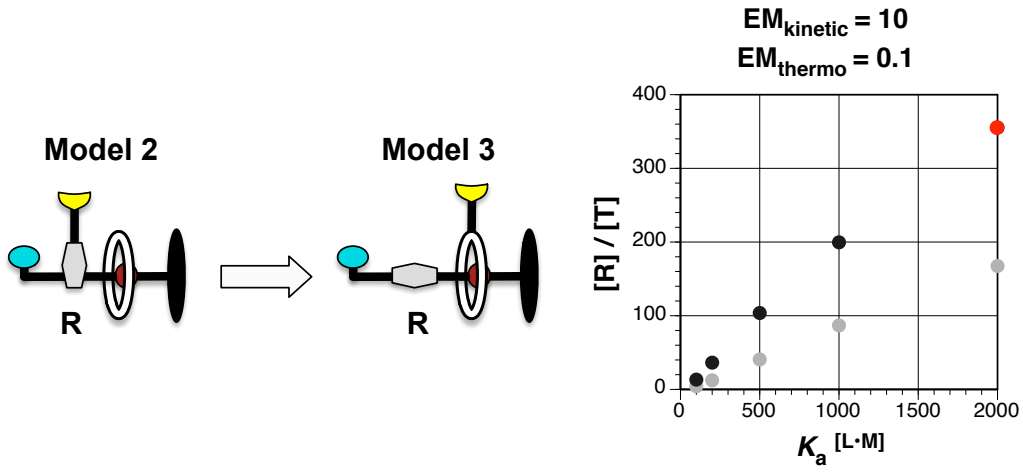


Figure 6.35 Design of a reaction network based on Model 3, where one recognition site required for self-replication is located on the macrocycle component can lead to a rotaxane–thread system that is highly selective for rotaxane formation, as evidenced by kinetic simulations investigating the effect of $K_a^{[\mathbf{L}\cdot\mathbf{M}]}$ on the ratio of [**R**]/[**T**] formed. The highest ratio of **R**]/[**T**] is highlighted in red. Grey and black circles represent ratios determined after four and eight hours, respectively.

CHAPTER 7

FUTURE WORK

The work presented in this thesis has examined a number of interconnected chemical systems based on small synthetic molecules. Using these networks, selectivity for one recognition-mediated reaction process over another was probed within various reaction environments. Namely, a network of competing replicators was examined initially using the established well-stirred batch reactor model, where all components are mixed homogeneously, and the outcome of replication depends on the kinetic and thermodynamic parameters governing the reaction processes, as well as the concentration of the reagents (**Chapter 3**). Later, the same interconnected network was examined within the environment of dynamically-exchanging components (aldehydes and nucleophiles), where the two replicators had to accomplish their synthesis using the building blocks distributed amongst the entire library (**Chapter 4**). The network was examined further under reaction-diffusion environment—far-from-equilibrium conditions where the outcome of the competition between two replicators depends also on the diffusion parameter (**Chapter 5**). Finally, the design for the more efficient replicator was exploited for the construction of a reaction network that integrates another, recognition-mediated processes that allowed this network to direct the assembly and replication of a [2]rotaxane.

In summary, the experimental work in this thesis focused on several phenomena in particular: selectivity in replicating networks, dynamic covalent systems coupled to irreversible self-replication processes, self-replication within reaction-diffusion environment and design of mechanically-interlocked architecture with the capacity to template its own synthesis. There are a number of experiments that could be used to probe further these phenomena, and they will be presented in the next sections.

7.1 Replicating systems under flow conditions

Throughout this thesis, the replicating networks are, with the exception of the reaction-diffusion environment probed in **Chapter 5**, examined the under well-stirred batch reactor (WSBR) conditions—*i.e.* a closed system environment. In this reaction format, no exchange of matter takes place with the outside world, which means that a replicator only has a finite amount of resources available for its formation. This limiting amount of reagents available means that kinetic selection is allowed to prevail, and the selectivity in a network of two competing recognition-mediated reactive processes remains limited.

A continuously stirred tank reactor (CSTR) model (**Figure 7.1a**) provides an environment for the examination of replicating systems where fresh unreacted starting materials are continuously, or at regular intervals, introduced into the tank. Unlike a reaction-diffusion system, the tank is continuously stirred, ensuring even mixing between the freshly added material and the contents already present. Simultaneously, a fraction of the material contained in the tank is removed in order to compensate for the quantity of the solution added, thereby avoiding overflow. In fact, the rate of material inflow should be equal to the rate of outflow, in order to maintain constant volume.

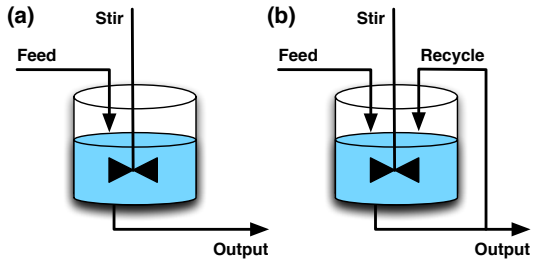


Figure 7.1 **(a)** A continuously stirred tank reactor (CSTR) in which reagents are introduced into a tank equipped with a stirrer. Simultaneously, material is also taken out of the tank in order to ensure steady flow and to avoid emptying or overfilling of the tank. CSTR shown in **(b)** includes an addition recycling loop, where the material output is used to instruct the reaction mixture by being introduced back into the reactor.

The reaction profile of a self-replicating system is such that in the absence of added preformed template, the reaction often exhibits an initial lag phase. Once the quantity of the formed template is sufficient for its assembly with the unreacted building blocks in to the ternary, catalytically-active complex, replication can proceed efficiently, until the reactants are depleted (*i.e.* the plateau phase). Throughout the duration of the reaction, therefore, self-replicating systems in a closed environment driven by kinetic selection (WSBR model) operate efficiently only for a limited amount of time. This limitation is reflected in the fact that the selectivity in a network containing more than one replicator, competing for a shared building block, is limited. A CSTR model, on the other hand, opens up the possibility of examining reaction networks in a situation

where the replicators can operate under optimum conditions throughout the duration of the experiment. Ultimately, such optimum conditions have the potential to allow the selectivity for the more efficient recognition-mediated reactive processes to increase over time. The extent to which the selectivity can differ relative to a simple WSBR environment is challenging to predict, and will require experimental testing assisted by kinetic simulations. The flow environment available for the study of replicating networks through the medium of CSTR is also more reminiscent of the conditions that a prebiotic chemical replicator might have experienced during the process of molecular evolution, particularly if the flow system is optimised to incorporate a dynamic pool of exchanging components as the feed-stock for the formation of the replicators.

Experimental investigation of selectivity in replicating networks will require development and optimisation of reaction conditions, namely the inflow and outflow rate, reaction concentration and temperature. Another parameter that could be exploited in the study of selectivity is the possibility of utilising the output from the reactor as an instructing material—*i.e.* a portion of the outflow is returned to the tank containing the reaction mixture (**Figure 7.1b**). The network of two replicators, **T1** and **T2**, examined in **Chapter 3** to **Chapter 5** presents an ideal platform for the investigation under CSTR conditions: the behaviour of both replicators has been well-established through the work undertaken in this thesis, both in isolation and within competition environment. Similarly, the ratio of the two replicators has been comprehensively analysed through various competition experiments, which means that it would be possible to compare the results obtained in a closed system to the selectivity determined under CSTR conditions, through the analysis of the material removed from the tank. As a result of the continued ability of the replicators to perform at their optimum efficiency, the environment of a flow system offers conditions where increased, if not complete selectivity between two competing template-mediated replicating processes can be observed.

7.2 Replicating systems under reaction-diffusion conditions: probing the diffusion parameter

The newly developed series of fluorescent tags explored in **Chapter 5** provide access to networks of self-replicating systems where each component is equipped with a distinct label, and its formation within an interconnected network can be monitored. The work presented in **Chapter 5** focused on a network of two replicators that exploit two auto- and one crosscatalytic pathway for their formation. In the future, this work exploiting reaction-diffusion conditions could be extended to include examination of replicating

networks with more varied catalytic relationships, as well as networks operating in the reciprocal sense.

The resolution of a replicating system of competing replicators examined within a reaction-diffusion media, be it self- or reciprocally replicating one, is governed by both its reactivity and diffusion. The fluorescent tags developed in this thesis provide a suitable strategy for analysis of the current two-replicator system in terms of its reactivity. As a result of the size similarity of the employed tags, however, exploration of the diffusion parameter space is not possible. A strategy can be envisaged, where an anthracene-based tag is redesigned (**Figure 7.2a**) so as to incorporate a dendrimeric appendage increasing progressively in size (**Figure 7.2b**).

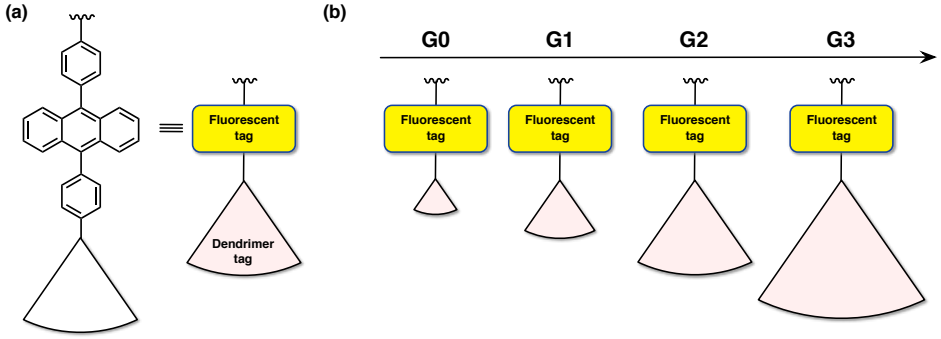


Figure 7.2 **(a)** Design of an anthracene-based fluorescent tag bearing a dendrimeric appendage and its simplified cartoon representation. **(b)** Attachment of increasingly large generations (**G0** → **G3**) of dendrimers to a fluorescent tag can permit exploration of the diffusion parameter space in the formation of propagating reaction-diffusion fronts in synthetic replicating systems.

In this manner, the diffusion parameter for any replicating system could be explored as a means of tuning and altering selectively. A comprehensive study of the diffusion and reactivity space would allow examination of how the interplay between these parameters affects the efficiency of wave propagation and selectivity within networks of competing replicators.

7.3 Reciprocally-replicating rotaxane network

Chapter 6 examined a reaction network exploiting orthogonal recognition processes to drive the replication and assembly of a [2]rotaxane according to the Model 2. Thus far, the experimental work reported in this thesis, as well as the previous work on integrating replication processes with rotaxane formation, exploited replication processes driven by self-complementary recognition processes only, *i.e.* those allowing the formation of self-replicating reaction networks only. A logical next step is to extend the lexicon of replicating [2]rotaxanes by examining the requirements for the implementation

of a network that integrates [2]rotaxane formation with reciprocal replication. For example, construction of a reciprocally-replicating (RR) rotaxane system based on Model 1 (**Figure 7.3**) can be envisaged from five building blocks (**L1**, **L2**, **S1**, **S2** and **M**) giving rise to two thread (**Figure 7.3a**) and two [2]rotaxane (**Figure 7.3b**) reciprocal replicators. Two core equilibrium processes between macrocycle **M** and linear components **L1** and **L2**, to give two pseudorotaxane complexes [**L1**·**M**] and [**L2**·**M**], mediate the formation of these replicators. A comprehensive set of kinetic experiments examining the reaction network both in the absence of added chemical input and in the presence of different preformed replicator templates, should allow the various catalytic and crosscatalytic pathways (**Figure 7.3c**) and their efficiency to be probed and established.

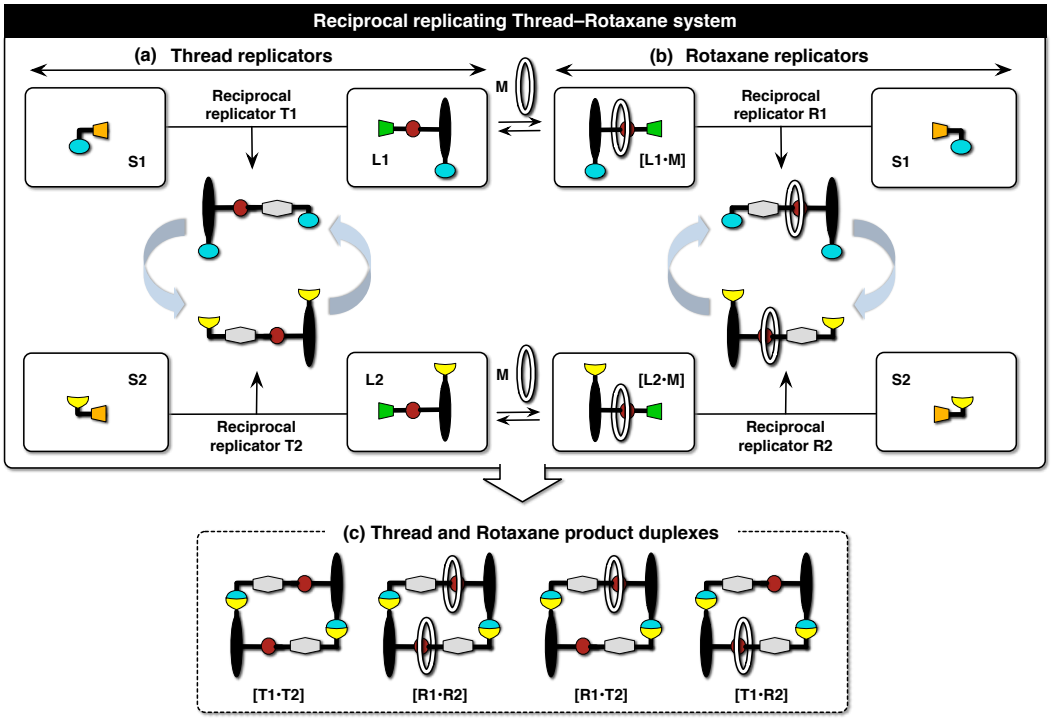


Figure 7.3 Design of a reciprocally-replicating rotaxane-thread network based on Model 1. The network is assembled from five building blocks: two linear components, **L1** and **L2**, two stoppers, **S1** and **S2**, and a macrocycle **M**. Reaction of these five components can produce **(a)** two thread reciprocal replicators and **(b)** two [2]rotaxane reciprocal replicators, with mutually complementary recognition sites. As a result of the complementary recognition sites, the network can form **(c)** four different template duplexes overall. Orange and green represent the reactive sites, whilst orange and blue denote recognition sites required for reciprocal-replication. Macrocycle binding sites are shown in red.

Whilst this suggested design of a RR network is based on Model 1, a reciprocally-replicating system integrating [2]rotaxane formation can be constructed also using Model 2 and Model 3. In fact, only comparison of the outcome and interplay of the various recognition and reaction processes operating in replicating systems based on

all three models can furnish a comprehensive picture of the requirements governing the behaviour of such hierarchical networks where replication processes and rotaxane assembly operate simultaneously.

7.3.1 Multicyclic rotaxane–thread network

An interesting scenario investigating the interplay between recognition and reactive processes operating simultaneously in a single system can be envisaged if a rotaxane–thread network performed in a minimal and reciprocal sense at the same time. Such highly interconnected network (**Figure 7.4**) incorporates an additional level of instruction. Comprised of the same five building blocks (**L1**, **L2**, **S1**, **S2** and **M**), this network can form both self- and reciprocal [2]rotaxane replicators (**Figure 7.4a**) as well as the macrocycle-lacking thread analogues (**Figure 7.4b**). Examination of how this multicyclic network would respond to the addition of various instructional preformed templates would likely reveal a vastly different behaviour, depending on the model used to construct the network.

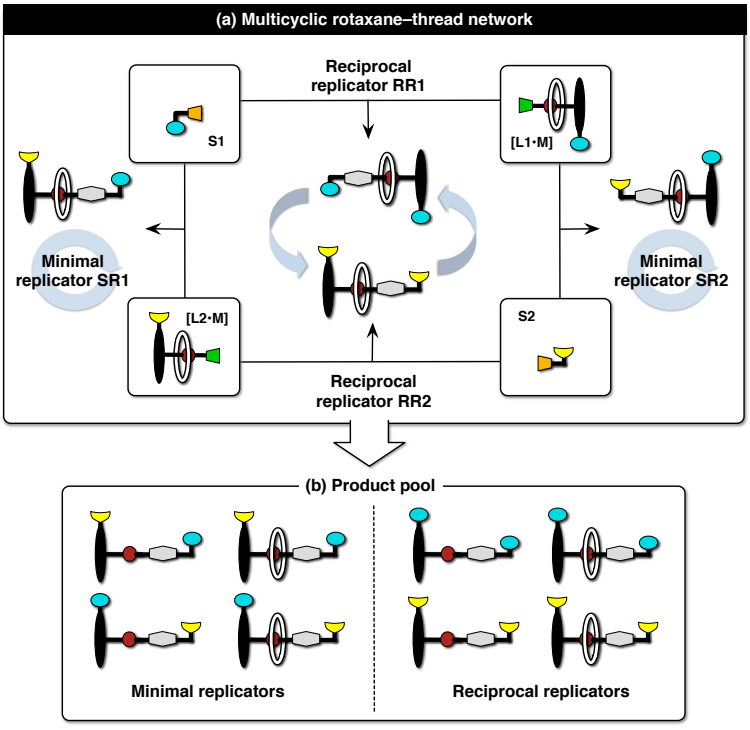


Figure 7.4 **(a)** Design of a multicyclic reaction network operating in the minimal and reciprocal sense simultaneously. The rotaxane-thread network is based on Model 1 and assembled from five building blocks: two linear components, **L1** and **L2**, two stoppers, **S1** and **S2** and a macrocycle **M**. **(b)** Reaction of these components can produce two thread and two [2]rotaxane reciprocal replicators, with mutually complementary recognition sites and the same number of minimal replicators, with self-complementary recognition sites. Orange and green shapes represent the reactive sites, whilst orange and blue denote recognition sites required for replication. Macrocycle binding sites are shown in red.

CHAPTER 8

CONCLUSIONS

Living organisms are undeniably complex—and we have only began to unravel and appreciate the level of their complexity. The widespread occurrence of living organisms around us leads to a question scientists are still trying to answer: what are the origins of this biological complexity, and indeed, life in general? The presence of living organisms means that, indisputably, the transition from simple molecular building blocks to what can be considered as life[a] must have occurred at least once. The identification of plausible steps that could explain the transition(s) from non-living matter to a world filled with life underlies the continuing endeavours[43] of the field of prebiotic chemistry. A central requirement of the origin of life is the emergence of molecules capable of self-replication—*i.e*. a process by which an entity is capable of templating its own synthesis. The reliance of template-direct processes for the copying and transmission of genetic material in current biological organisms is well-established. Systems chemistry complements prebiotic chemistry research by investigating chemical networks, in which the connections between individual molecules allow complex, dynamic system-level behaviour to emerge.

The focus of the work presented in this thesis has been on utilising the wide scope afforded by synthetic chemistry and the molecular recognition toolkit developed by the field of supramolecular chemistry, to engineer increasingly complex reaction networks that are constructed from simple components that possess the capacity to react and interact. The research in the field of prebiotic chemistry is driven by the desire to establish plausible routes to prebiotically-relevant molecules, *i.e*. molecules that are likely to have played a role in the process of chemical evolution, such as nucleic acid analogues,

[a]While our understanding and agreement as to what may constitute life remains a source of significant debate, NASA has provided[38] the generally most accepted definition of life, where "a self-sustaining chemical system capable of undergoing Darwinian evolution" is considered to be living.

RNA, fatty acid chains as membrane components and amino acids. In contrast, the bottom-up approach epitomised by systems chemistry moves beyond the restrictions imposed by studying molecules with a specific relevance to current biology—focusing instead on the study of complex phenomena, in particular self-replication, using synthetic systems as models. Natural systems often exhibit a level of complexity that makes identifying and decoupling all of the individual interactions and reaction processes that occur simultaneously extremely challenging, if not impossible. Inspired by natural systems, systems chemistry strives to maintain structural and interactional simplicity in the network components, in order to design systems with well-defined chemistries and interactions, which can be analysed and characterised experimentally. In the simplest terms, a self-replicating system is constructed from two molecules, where the product of their reaction is a catalyst for its own formation—giving rise to a non-linear reaction profile, thus allowing construction of systems responsive to feedback, *i.e*. addition of preformed template as chemical input. Experimental systems based on the simplest model of self-replication have demonstrated that enzyme-independent, template-driven replication is possible in synthetic systems based on oligonucleotides, peptides and small molecules—without the need for the complex enzymatic machinery required by the modern cell.

A primitive model of competition in biological systems can be constructed using just three building blocks, which react to produce a network of two interconnected replicators. In the work presented in this thesis, such a reaction network was assembled by the reaction of two simple small-molecule components with a shared building block. The initial task of establishing the catalytic efficiencies of the various auto- and crosscatalytic pathways available to this designed system was accomplished through a comprehensive set of kinetic experiments, exploiting the ability of replicating systems to respond to addition of preformed template. Small variations in the structural features of the employed components gave rise to a two-replicator network where both autocatalytic pathways are perform efficiently, but only a single crosscatalytic cycle is active. The comprehensive kinetic analyses revealed that the replicator, which is capable of forming less strongly-bound template duplexes, is also the one that is capable of more efficient recognition-mediated covalent capture. As a result of the weaker strength of the interaction between the two the complementary recognition elements, required for replication, this same replicator exhibited also a diminished ability to assemble the reaction components in to a catalytically-active ternary complex. Nevertheless, the advantage afforded to this replicator as a result of the higher catalytic turnover (*i.e*. lower product inhibition) allowed it to perform more efficiently in isolation. By contrast, the second replicator with a higher duplex association constant and a stronger association

between the two recognition elements turned out to be less efficient at utilising the building blocks for its own formation. In this case, once the product was formed, the higher strength of the template duplex resulted in a decreased proportion of the free catalytically-active template, that is required for efficient template-directed pathway.

The analysis of this three-component network under competition conditions—*i.e.* an environment where only a limited amount of the shared building block, which is necessary for the construction of both replicators, was available—showed that even a thorough characterisation of all the individual reaction pathways can sometimes be insufficient for predicting the behaviour of the replicators when they interact with each other within a single system. In this case, the interplay of recognition and reaction processes allowed the replicator that was found to be weaker in isolation to become the dominant species in the reaction mixture. This outcome illustrates clearly that higher selectivity for one replicator over another, which is analogous to higher fitness[b] in biological systems, depends on both the environment it is found within, as well as other species present at the same time, all competing for a limited stock of shared resources. The attempts to bias the competition between one replicator over another using preformed templates as instruction achieved limited success. The competition in this two-replicator network was examined under well-stirred batch reactor conditions—a closed reaction environment that imposes a limit on the level of selectivity achievable in a system of two interconnected, recognition-mediated reaction processes. This kinetic barrier results in the erosion of the initial imbalance between the two replicators: the exhaustion of the building blocks in the system limits the efficiency of the replication processes, thereby stopping the more efficient replicator from gaining a significant advantage.

Examination of replicators in isolation within a closed reaction environment presents a unique opportunity to analyse and establish the recognition and reaction processes governing the formation of these species in the absence of other competing processes. Nevertheless, the reaction and recognition processes within complex networks found in the real world do not operate in isolation—nor do they operate using fully preformed components. A more complex environment of dynamically-exchanging components for the examination of replicating systems can be constructed using dynamic covalent chemistry (DCC)—a tool that combines the general combinatorial approach with the benefits of reversible bond formation. Moving away from the reaction environment governed purely by kinetic selection, the two-replicator network can be examined under conditions where the shared building block required for the formation of both

[b] In biology, fitness can be used to describe the ability of an organism to transfer information to its progeny

replicators is not yet synthesised at the onset of the reaction. Within the environment of a dynamic library, the replicators have to exploit the molecular recognition engineered in to the components themselves, in order to accomplish their formation from the building blocks distributed across the entire library. In the absence of instruction, *i.e.* any components which can react with those present within the library, a dynamic library reaches a distribution that reflects the most thermodynamically-favourable position. However, if a component capable of transforming the library through irreversible kinetic processes is added, the library is driven away from its equilibrium position in order to replenish the library components that are being consumed at the fastest rate—*i.e.* the library components necessary for the formation of the replicators. The system-level behaviour observed as a result of the library-equilibration taking place is the increase in the concentration of the components not required for the formation of the components formed through the recognition-mediated processes. The fate of the library and the resolution of the replicator network were also examined in the presence of instructing preformed template. Whilst the library composition revealed a marked preference for the products formed through the template-mediated pathways, the selectivity for one replicator over another remained essentially unaffected by the dynamic environment under the experimental conditions employed, and instead it was governed by kinetic selection.

Using the experimental system as a baseline, the rules governing the selectivity for one replicator over another in a system where covalent and non-covalent selection processes operate simultaneously were probed further using kinetic simulations. The simulations undertaken in this thesis show that two parameters in particular can significantly affect the efficiency of replicating systems: catalytic efficiency and the stability of the template duplex relative to the stability of the ternary complex. In particular, highly efficient replication within both a kinetically-driven environment and that governed by dynamic selection require: (i) a significantly higher efficiency of the template-mediated pathway relative to the bimolecular reaction (high $EM_{kinetic}$), and (ii) a lower stability of the template duplex relative to the stability of the ternary complex (low EM_{thermo}). Additionally, the simulations showed that variation in the concentration of the reactants can alter the selectivity significantly—in particular, examination of the two replicators under competition conditions where the concentration is below the K_d for one of the recognition-mediated processes can provide a dramatic increase in selectivity. Nevertheless, the increase in selectivity at low concentration of reactants is often accompanied by extremely low conversions and rates of reaction. Whilst the slow conversion achieved at low reaction concentrations might make analysis challenging, it should be remembered that the formation of a molecule capable of templating its own synthesis on the prebiotic

earth also required a substantial length of time—a process that most likely involved development of a boundary, which separated such replicator from the rest of the world that enabled it to sustain itself. Further simulations examining the possibility of increasing the selectivity for one replicator over another by addition of preformed template corroborated the observation determined experimentally—the selection is limited and the advantage afforded to a replicator by the addition of preformed template is eroded over time as a result of the closed nature of reaction environment.

Life is an emergent phenomenon arising in a complex network of interconnected components and processes, that is characterised[24,39,41] by a state of far-from-equilibrium dynamic kinetic stability. The structural features and uniformity of the current genetic material suggest that the modern living systems all emerged from a single Last Universal Common Ancestor (LUCA), the emergence of which also required conditions that are far-from-equilibrium. In addition to emergence of life, it has been well-established that far-from-equilibrium environment can allow complex behaviour and phenomena such as spatio-temporal patterns and oscillations to emerge. The closed reaction environment employed for the study of the interconnected network of replicators has been shown to result in a mixture of components—*i.e.* conditions where both replicators are able[71,320,321] to coexist. In the presence of instructing template, the selectivity for one replicator could be altered, but only to a limited extent. In order to move beyond the boundary imposed on selectivity by the closed reaction environment, it is necessary to explore 'open', non-homogeneous reaction conditions that are far-from-equilibrium—an environment that is so prevalent in nature. The work presented in this thesis explored the possibility of using the environment of a reaction-diffusion front as a model of open, far-from-equilibrium conditions for the study of networks of replicators. In the reaction-diffusion environment, the autocatalysis is coupled to diffusion, allowing a solution of preformed template added at a specific location to a mixture of unreacted components to establish a propagating reaction-diffusion front, driven by the resulting concentration gradient. Under these conditions, the initiated reaction-diffusion front ensures that the replicator template will encounter unreacted building blocks as it propagates, permitting the replicator to operate at its optimum efficiency throughout the reaction—a situation where the replication processes do not become self-inhibiting and an initial imbalance in selectivity can propagate.

Experimental analysis of the competition in a network of two replicators within a reaction-diffusion environment revealed selectivity higher than that observed in a template-instructed experiment performed under closed reaction conditions, suggesting that the far-from-equilibrium conditions allow the system to move beyond the limit imposed by kinetic selection within a closed reaction environment. The increased

selectivity demonstrated in the proof-of-principle experimental studies agree with the theoretical predictions put forward[239] by Showalter and co-workers, who simulated a network of two competing autocatalytic processes within reaction-diffusion media, showing that complete selectivity for one replicator over another is possible. These simulations, however, examined a system where the two autocatalytic products could not be formed through bimolecular reactions—instead, the formation of the autocatalytic products was modelled as a simple second order reaction with respect to the reaction components. In a network of self-replicators, which first need to be formed through the bimolecular pathways before the template-mediated cycles can operate, it will only be possible to observe a near-complete selectivity for one replicator over another. In the future, simulations investigating competing replication processes based on the minimal model of self-replication will hopefully complement the experimental studies presented in this thesis, and help answer the questions that currently remain unanswered. Namely, construction of a general simulation model for the analysis of networks of competing replicators could probe (i) whether the catalytically-active ternary complexes involving template become saturated at the edge of the propagating reaction-diffusion front, (ii) what are the limits on reaction and diffusion parameters that can give rise to propagating reaction-diffusion fronts, and finally, (iii) to examine how the experimental outcomes compare to those determined through simulations. Ultimately, simulations of replicating systems under both reaction-diffusion environment could be extended to systems operating under flow conditions, investigating also networks operating in the reciprocal sense.

The studies undertaken on a network of competing replicators in this thesis illustrate clearly that the resolution of a reaction network is tied closely to the selection processes operating in the network, as well as the reaction environment under which it is examined. An area of replicating systems that remains vastly under-explored is the study of replicating systems in the presence of additional non-covalent selection processes. In this thesis, such a system was studied by extending the design of an efficient self-replicating system to integrate another recognition-mediated non-covalent selection algorithm, in addition to the recognition-mediated processes required for the covalent capture, driving replication—specifically, a recognition process required for the assembly of a rotaxane mechanically-interlocked architecture within the same system. Several ways of organising the recognition and reactive sites exist in a system incorporating replication processes with rotaxane assembly. The experimental investigations of the designed system showed that the non-covalent processes required for the assembly of a rotaxane have little effect on the efficiency of the replication processes if they are located at the periphery of the covalent framework, instead giving rise to a system

where the rotaxane and the non-mechanically-interlocked analogue are equally efficient as catalytic templates in both auto- and crosscatalytic pathways. At the same time, the study revealed that as a result of an unforeseen interplay between the recognition and reactive processes operating in the system in parallel, a selection between the two catalytic templates is inadvertently biased towards the formation of the non-interlocked product. Kinetic simulations revealed that if the network is redesigned in such a way that the second recognition-mediated interaction is directly integrated with the replication processes—*i.e*. a situation where both processes are required for the template-directed self-replication, the selectivity of the reaction network for the rotaxane architecture over the non-interlocked product increases dramatically.

Probing the requirements that allow and govern molecular replication is crucial for our understanding of the processes governing chemical evolution—*i.e*. the gradual transition culminating in the emergence of a living world from a non-living one. To date, self-replication has been demonstrated experimentally in systems constructed from a variety of building blocks, ranging from more prebiotically-relevant oligonucleotide and peptide frameworks to completely synthetic systems. With the advances in analytic tools and methods, it is now possible to analyse and follow the behaviour of complex chemical networks and systems where multiple catalytic pathways perform simultaneously—moving beyond the study of molecular matter in isolation. The research described in this thesis illustrated that through careful engineering of the reaction components of chemical networks, construction of instructable networks of replicators, where auto- and crosscatalytic pathways operate in parallel, is possible. Utilising only the instructions encoded within the constitutional frameworks of their components, these networks can direct their own formation, exhibiting system-level behaviour that goes beyond the properties expressed by their components in isolation. For the firs time, through incorporation of an additional, appropriately designed non-covalent selection algorithm, a reaction network can exploit replication processes to drive the assembly and replication of a mechanically-interlocked architecture. Exciting opportunities for the investigations of replicating systems are emerging both in the form of DCC and far-from-equilibrium reaction environments. The proof-of-principle realisation of reaction-diffusion fronts driven by small-molecule based synthetic replicators in this thesis provides a platform for the investigation of networks of competing replicators under conditions that are far-from-equilibrium—environment where selectivity for one replicator over another is not limited by the constraints imposed on selectivity by closed reaction conditions. Nevertheless, whilst the drive towards the examination of networks of self-replicators under more complex conditions will likely remain, so will the need for

analysis of these systems under simple closed system conditions where the individual reactions and interactions can be decoupled, characterised thoroughly and analysed.

CHAPTER 9

EXPERIMENTAL PROCEDURES

9.1 General experimental procedures

All chemicals and solvents were purchased from Sigma-Aldrich Company Ltd., Alfa Aesar UK, Apollo Scientific Ltd., Fisher Scientific UK Ltd., TCI UK Ltd., VWR International Ltd. or Acros UK, and were used as delivered, unless otherwise stated. Where appropriate, all non-aquaeous reactions were carried out under inert atmosphere. Dry THF, PhMe and CH_2Cl_2 obtained using an MBraun MS SPS-800 solvent purification system, where solvents were dried by passage through filter columns and dispensed under an atmosphere of argon atmosphere. Thin layer chromatography (TLC) analysis was performed using plastic 0.20 mm MACHEREY-NAGEL GmbH & Co. POLYGRAM SIL G/UV254 plates or aluminium TLC Silica gel 60 F254 plates. Plates were air-dried and visualised under UV lamp (254 nm or 366 nm). Alternatively, compounds containing primary and secondary amines were visualised by dipping in ninhydrin stain, followed by heating. Flash column chromatography was performed using SillaFlash® P60 silica gel ($230-400$ mesh, Sillicycle). $PdCl_2(PPh_3)_2$ and $Pd(PPh_3)_4$ catalysts were prepared according to standard[322] synthetic procedures. Samples of thread **T** and both rotaxanes, **GM-R** and **PM-R**, were purified using reverse phase column chromatography using C18 ISOLUTE columns (10 and 25 g) purchased from Biotage. Melting points were determined using Stuart SMP30 melting point apparatus or an Electrothermal 9200 melting point apparatus. Electrospray ionisation spectra (ESI) were performed on a Micromass LCT spectrometer, whilst chemical ionisation (CI) spectra were obtained using Micromass GCT spectrometer, operating in positive or negative mode from solutions of MeOH, MeCN or H_2O. Mass-to-charge ratios (*m/z*) in reported mass spectra are provided in Daltons.

9.2 General NMR spectroscopy procedures

Deuterated solvent was used as the lock and the residual solvent as the internal reference in all cases. ^{1}H, ^{13}C and ^{19}F$\{^{1}$H$\}$ spectra were analysed using iNMR software (Mestrelab Research) or MestReNova (Mestrelab Research).

9.2.1 ^{1}H NMR spectroscopy

^{1}H NMR spectra were recorded on either a Bruker Avance 300 (300.1 MHz), Bruker Avance II 400 (400.1 MHz), Bruker Avance 400 (400.3 MHz), Bruker Avance III 500 (499.9 MHz) or a Bruker Avance III-HD 500 (500.1 MHz) spectrometer using the deuterated solvent as the lock. The chemical shift information (δ_{H}) for each resonance signal is given in units of parts per million (ppm) relative to centre of solvent resonance ($CDCl_3$: 7.26 ppm, DMSO-d_6: 2.50 ppm). The number of protons for a reported resonance signal is indicated as nH based on individual integral values and multiplicity denoted by the symbol in parentheses. Coupling constants (J) are determined using iNMR and are quoted to the nearest 0.1 Hz.

9.2.2 ^{13}C NMR spectroscopy

^{13}C NMR spectra were recorded on either a Bruker Avance 300 (75.5 MHz), a Bruker Avance II 400 (100.6 MHz), Bruker Avance 400 (100.7 MHz), Bruker Avance 500 (125.7 MHz) or a Bruker Avance III-HD 500 (125.8 MHz) spectrometer using the deuterated solvent as the lock. DEPTQ pulse sequences with broadband proton decoupling were employed during spectra recording. The chemical shift information (δ_{C}) for each resonance signal is given in units of parts per million (ppm) relative to the centre of solvent resonance ($CDCl_3$: 77.16 ppm, DMSO-d_6: 39.52 ppm). All signals are singlets unless stated otherwise.

9.2.3 ^{19}F$\{^{1}$H$\}$ NMR spectroscopy

^{19}F$\{^{1}$H$\}$ NMR spectra were recorded on a Bruker Avance 300 (282.4 MHz), Bruker Avance II 400 (376.5 MHz), Bruker Avance 400 (376.6 MHz), Bruker Avance 500 (470.3 MHz or 470.4 MHz) or a a Bruker Avance III-HD 500 (470.5 MHz) spectrometer using the deuterated solvent as the lock. Spectra were recorded using broadband proton decoupling pulse sequences. The chemical shift information (δ_{F}) for each resonance signal is given in units of parts per million (ppm) relative to CCl_3F (CCl_3F: 0.00 ppm) or 1-bromo-2-fluoro-4-nitrobenzene ($C_6H_3BrFNO_2$: 102.93 ppm (CD_2Cl_2) and 101.75 ppm ($CDCl_3$))

9.3 T1–T2 network

9.3.1 Kinetic measurements and deconvolution

Kinetic experiments described in **Chapter 3** were analysed by ^{1}H NMR (499.9 MHz) and by ^{19}F{^{1}H} NMR spectroscopy (470.3 MHz). Reagents were weighed out using a Sartorial BP211D balance ($\pm$ 0.01 mg) and dissolved in fresh solvent ($CDCl_3$), 99.8 atom % D) in 1 mL glass vial. The volume of solvent was measured using 1 mL Hamilton gastight syringes. The contents were sonicated and 0.9 mL of the volume was transferred into a 5 mm NMR tube (Wilmad 528PP). The NMR tube was equipped with a polyethylene pressure cap to prevent solvent evaporation. The sample was cooled on ice before being inserted into a Bruker Avance NMR spectrometer, regulated at the desired temperature (5 °C).

9.3.2 ^{1}H NMR spectroscopy

499.9 MHz ^{1}H NMR spectra were automatically acquired at a desired time interval for a period of 15 to 22 hours. Analysis and deconvolution of the collected data was performed using iNMR software (Mestrelab Research). Reaction progress was determined by monitoring of the disappearance of resonance specific for maleimide CH=CH protons, and the appearance of signals characteristic for the protons on the isoxazolidine ring system of the 1, 3-dipolar cycloaddition products (**Figure 9.1**).

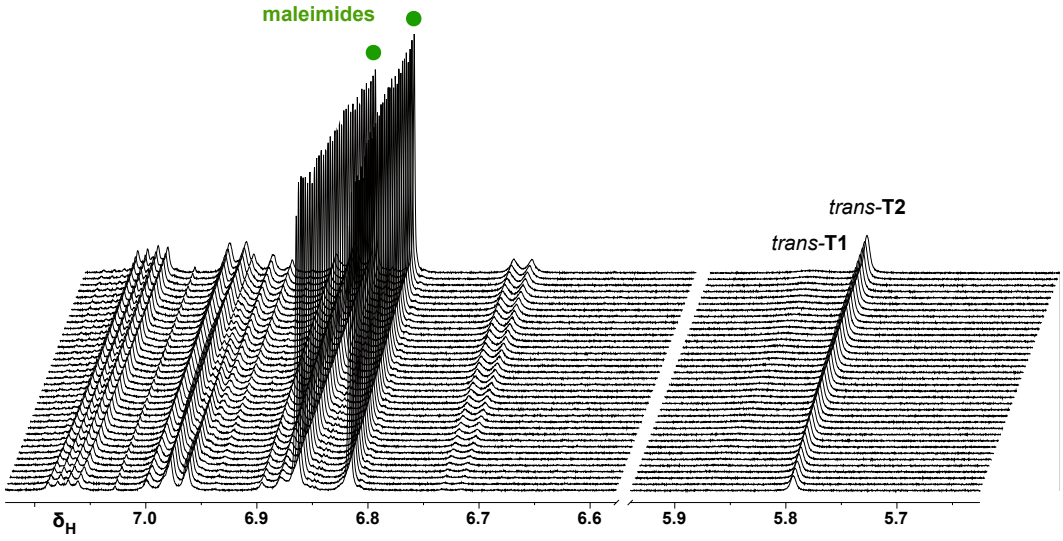


Figure 9.1 Partial plot of arrayed ^{1}H NMR spectra (499.9 MHz) recorded during the kinetic analysis of **M1** and **M2** reaction with $\mathbf{N^F}$ (all components at 5 mM, $CDCl_3$, 5 °C), instructed with *trans*-**T2** (1 mM). Important changes in the arrayed spectra over time are the appearance of *trans* resonances arising from the **T1** and **T2** cycloadducts at *ca.* 5.8 and 5.9 ppm and disappearance of maleimide protons at *ca.* 6.9 ppm.

Using ^{1}H NMR, it was possible to distinguish between the *trans* and *cis* relative configuration of the template since relative orientation of the three protons on each cycloadduct are different. For the *trans* isomer shown in **Figure 9.2**, the dihedral angle between H^1 and H^3 is approximately 90° and the Karplus equation predicts[323,324] the coupling constant to be 0 Hz (singlet). The dihedral angle between the H^2 and H^3 protons is close to 20°, and the coupling constant is expected to be around 8 Hz for both (**Figure 9.2**). Conversely, the dihedral angles in the *cis* product are all around 20°, giving rise to two doublets (8 Hz) and a doublet of doublets (an apparent triplet).

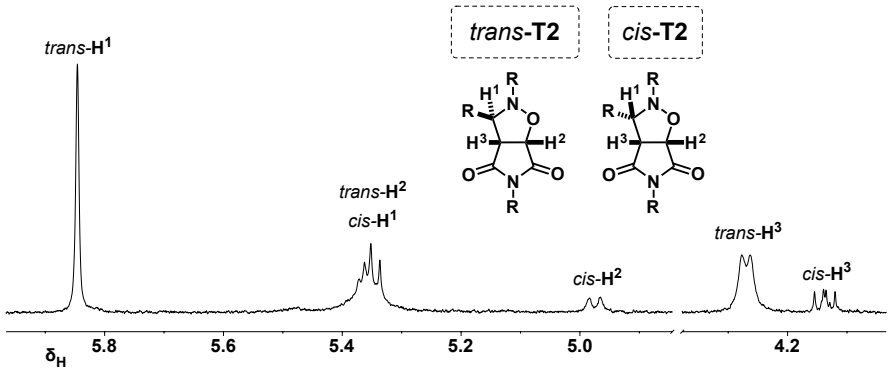


Figure 9.2 Partial 499.9 MHz ^{1}H NMR spectrum of a kinetic experiment between **M2** and $\mathbf{N^F}$ to give both *trans*-**T2** and *cis*-**T2** diastereoisomers. Splitting patterns typical for *trans* cycloadduct: s, d, d and for *cis* cycloadduct: d, d and dd are usually observed. In this experiment, doublets arising from the *cis* and *trans* overlap.

9.3.3 ^{19}F$\{^{1}$H$\}$ NMR spectroscopy

Reaction progress was also monitored by 470.3 MHz ^{19}F$\{^{1}$H$\}$ NMR spectroscopy, over 15 to 22 hours, employing a 10 to 15 second delay time between individual scans. Throughout deconvolution, particular focus was on the fluorine resonances associated with the the nitrone and both *cis* and *trans* cycloadduct products (**Figure 9.3**).

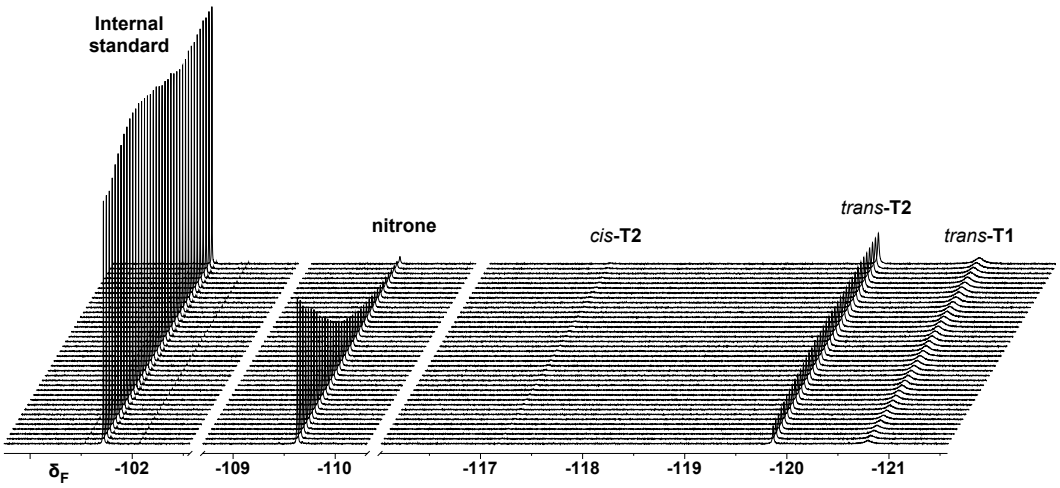


Figure 9.3 Partial plot of arrayed $^{19}F\{^{1}H\}$ NMR spectra (470.3 MHz) recorded during the kinetic analysis of **M1** and **M2** reaction with $\mathbf{N^F}$ (all components at 5 mM, $CDCl_3$, 5 °C), instructed with *trans*-**T1** and *trans*-**T2** (1 mM each). Important changes in the arrayed spectra over time are the appearance of *trans* resonances arising from the **T1** and **T2** cycloadducts at *ca.* −119.8 and −120.8 ppm and disappearance of nitrone resonance at *ca.* −109.6 ppm.

9.3.4 Semi-automatic deconvolution

Data recorded by ^{1}H and $^{19}F\{^{1}H\}$ NMR kinetic experiments were analysed using the semi-automatic deconvolution feature in the iNMR software package (Mestrelab Research). Using an initialisation script file, together with a script adapted for each specific experiment (both written in the scripting language Lua), the software package used an arrayed FID file (*e.g.* **Figure 9.3**) of all the desired spectra to perform a least squares optimisation to determine the area of selected peaks (**Figure 9.4a**). Residual error in the calculation process can be determined using the toggle function (**Figure 9.4b**). Exact concentrations for reactants and products were determined relative the area of an internal standard at a known concentration. 1-Bromo-2-fluoro-4-nitrobenzene was used for the kinetic experiments described in **Chapter 3** and for analysis of dynamic covalent libraries presented in **Chapter 4**. The resulting concentration *vs* time data was plotted using proFit software package (Quantum Soft). An example script can be found in the **Appendix A1**.

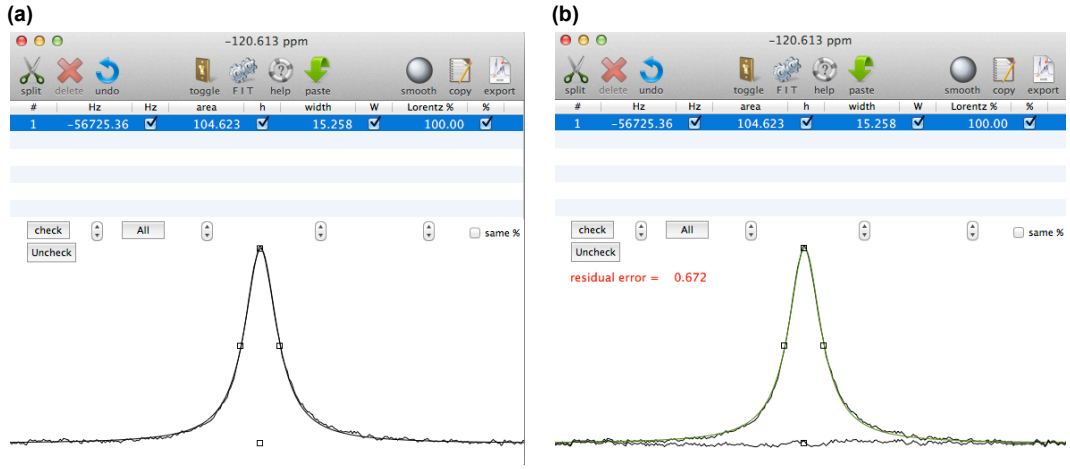


Figure 9.4 Example of the iNMR semi-automatic deconvolution window (**T1** signal in a 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectrum). **(a)** Area of a singlet peak is determined using the least-square fitting method. **(b)** Residual fitting error is displayed through the toggle function.

9.3.5 Error estimation in NMR kinetic experiments on T1 and T2 replicators

A signal-to-noise measurement was undertaken on several spectra from the kinetic experiments on the **T1**–**T2** network (in **Chapter 3**) using the *.sino* function in TopSpin (Bruker BioSpin, Version 3.5 pl5, 2016). The signal region in the $^{19}F\{^{1}H\}$ NMR spectra was selected to contain the resonance associated with the 1-bromo-2-fluoro-4-nitrobenzene internal standard. The function calculated the signal-to-noise ratio (S/N) for this peak, which was used as a reference (of known concentration) for determining the concentration of all the other species in the reaction mixture. The S/N ratio was typically calculated to be between 740 and 1260 and the concentration of the internal standard in the kinetic experiments was in the range 4 to 7 mM, making the lowest detectable concentration approximately 5 to 10 µM (an error in concentration determination of $\pm$ 0.02 mM).

9.3.6 Kinetic fitting and simulations

Kinetic fitting for data obtained from kinetic experiments and simulations were performed using SimFit software package (Version 32, kindly provided by Günther von Kiedrowski, Ruhr-Uversität, Bochum, 2008). In all cases, a kinetic model describing all the possible interactions in a system was constructed. SimFit converts the model into a series of rate equations and uses a Simplex algorithm to fit the kinetic parameters to the experimental data (generally, the kinetic parameters are allowed to vary by a factor of $\pm$ 1000). The difference between the fitted and experimental data is reflected by the residual % RMS error. The obtained rate constants k_{uni} and k_{bi} were used to calculate

the kinetic effective molarity ($EM_{kinetic}$) and the free energy of connection (ΔG^S) for each replicator.

Once the kinetic parameters were available for each autocatalytic and crosscatalytic pathway, simulation models were constructed employing these values. The kinetic parameters were not allowed to vary, and the reaction time and the initial concentration of the starting components and the template added were the only variables that could be altered. Example fitting and simulation scripts can be found in the **Appendix A2**.

9.3.7 Heteroduplex NMR study

Chapter 3 examined the changes in the chemical shifts of resonances associated with the *trans*-**T1** and *trans*-**T2** products by ^{1}H and ^{19}F$\{^1$H$\}$ NMR spectroscopy. The reaction samples, containing the two templates at a combined concentration of 10 mM (whilst the ratio of [**T1**]/[**T2**] was varied) were prepared by mixing of the nitrone $\mathbf{N^F}$ and maleimides (**M1**, **M2** or both) at the desired concentration (*e.g.* to prepare the 1 : 1 mixture of **T1** and **T2**, $\mathbf{N^F}$ was added at 10 mM whilst both maleimides at 5 mM) and left to react in $CDCl_3$ at 5 °C until complete conversion to products was determined by NMR analysis.

9.4 Dynamic covalent libraries

9.4.1 Preparation and analysis of DCLs

Chapter 4 described analysis of thirteen different DCL samples. The method of assembling the necessary components for each condition is described in **Figure 9.5**. To begin with, a stock solution of the reaction solvent was prepared by stirring CD_2Cl_2 with excess *p*TSA monohydrate under argon for 30 minutes, followed by filtering of the excess *p*TSA monohydrate. In the next step, a known amount of 1-bromo-2-fluoro-4-nitrobenzne as an internal standard was introduced into the solution (**Figure 9.5a**).

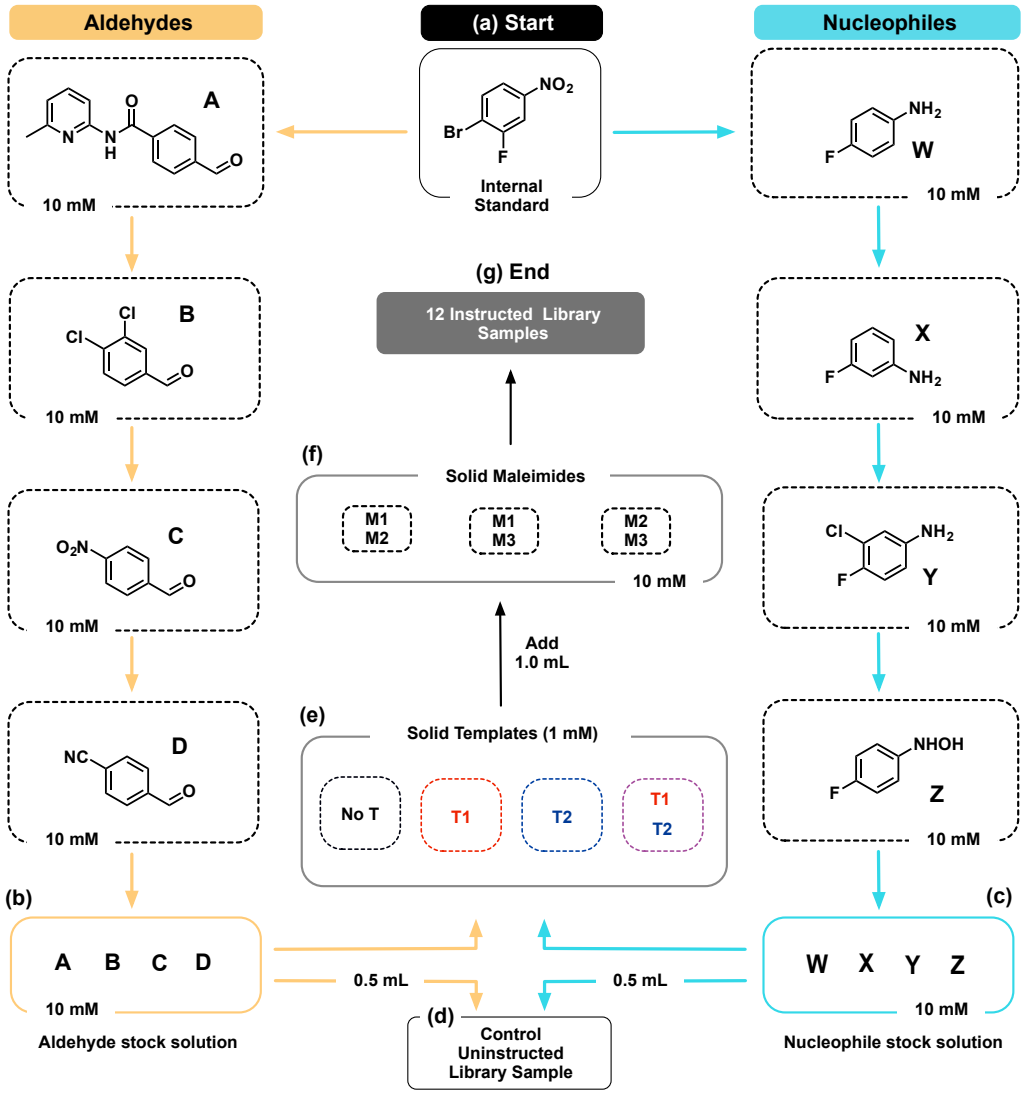


Figure 9.5 Flow chart illustrating the preparation of DCL samples for $^{19}F\{^{1}H\}$ NMR analysis. Aldehyde and Nucleophile stock solutions were prepared using CD_2Cl_2 saturdated with *p*TSA monohydrate with an internal standard at a known concentration. The two stock solutions were used to prepare a control library sample and were added to a vial containing preformed templates (solid). The afforded mixtures were used to dissolve the required combination of solid maleimides.

All library components were weighed out into individual glass vials using a Sartorial BP211D balance ($\pm$ 0.01 mg). Hamilton gastight syringes (1 mL and 5 mL) were employed for dispensing required volume of stock solutions. The internal standard stock solution was used to prepare the required amount of 10 mM solution of aldehyde **A**. This solution was added to a solid sample of aldehyde **B**, such that both components were at 10 mM. The process was repeated with aldehydes **C** and **D**, affording a stock solution of the four aldehydes at 10 mM concentration (**Figure 9.5b**). In a similar fashion, the prepared stock solution of internal standard was used to prepare a 10 mM stock solution of the four nucleophiles (**Figure 9.5c**).

With these two stock solutions now available, it was possible to proceed to preparation of the library NMR samples. To prepare the control library sample, incorporating no maleimides or templates, 0.5 mL of each stock solution was added to a glass vial and the contents were sonicated (**Figure 9.5d**). The sample was then transferred into a 5 mm NMR tube (Wilmad 528PP). A polyethylene pressure cap was applied to the sample to prevent solvent evaporation.

The typical procedure for preparing one of the 12 remaining libraries is illustrated in **Figure 9.5** and can be described as follows. Template stock solutions (no T, **T1**, **T2**, and both templates at 1 mM) were prepared by dissolving appropriate amounts of solid templates in 1 : 1 volumes of aldehyde and nucleophile stock solutions (**Figure 9.5e**). The four stock solutions were used immediately in the next step. Combinations of two maleimides were pre-weighed into a single 1 mL glass vial, such that the addition of 1 mL of each template stock solution afforded the desired library sample with both maleimides at 10 mM concentration (**Figure 9.5f**). Following mixing and sonication, the contents were transferred into an NMR tube and pressure cap was applied (**Figure 9.5g**). The 12 freshly prepared instructed library samples and one control library sample were left to equilibrate and evolve in a thermally controlled water bath (at 5 °C). Distribution of the exchange and product pool components was determined by 282.4 MHz or 470.3 MHz $^{19}F\{^{1}H\}$ NMR spectroscopy after a period of two and seven days.

The exact chemical shifts (282.4 MHz, CD_2Cl_2 saturdated with *p*TSA monohydrate, 5 °C), determined relative to 1-bormo-2-fluoro-4-nitrobenzene (-102.93), and the relaxation times (470.3 MHz, CD_2Cl_2 saturated with *p*TSA monohydrate, RT) determined for the DCL components are summarised in **Table 9.1**.

Table 9.1 List of $^{19}F\{^{1}H\}$ NMR signals (282.4 MHz, CD_2Cl_2 saturdated with *p*TSA monohydrate, 5 °C and relaxation delay times (470.3 MHz, CD_2Cl_2 saturated with *p*TSA monohydrate, RT) for the library components. The *cis*-**T2** was the only *cis* product assigned. Values marked – have not been determined. ⋆ Internal standard denotes 1-bromo-2-fluoro-4-nitrobenzene. ⋆⋆ Chemical shifts can vary as a results of formation of recognition-mediated complexes and duplexes.

Compound	Chemical shift (ppm)	Relaxation time (s)	Compound	Chemical shift (ppm)	Relaxation time (s)
internal standard*	−102.93	1.18	**AY**	−119.71	3.02
CZ	−110.05	–	*trans*-**T2C**	−120.14	1.01
DZ	−110.26	–	*trans*-**T2D**	−120.24	1.00
BZ	−110.60	–	*trans*-**T2B**	−120.32	1.03
AZ	−110.66	–	*trans*-**T2****	−120.64	0.71
CX	−112.91	2.64	*trans*-**T1C**	−120.78	1.30
DX	−112.98	2.54	*trans*-**T1D**	−120.89	1.26
BX	−113.07	2.48	*trans*-**T1B**	−120.95	1.23
AX	−113.02	1.52	*trans*-**T1****	−121.43	0.81
X	−114.04	4.14	*trans*-**T3C**	−120.80	–
CW	−116.23	3.99	*trans*-**T3D**	−120.90	–
DW	−116.52	4.00	*trans*-**T3B**	−120.97	–
BW	−116.99	3.79	*trans*-**T3A**	−121.39	–
AW	−117.09	3.42	**Z**	−123.15	–
CY	−118.82	3.68	**W**	−126.68	–
DY	−119.11	3.54	**Y**	−131.19	4.78
BY	−119.58	3.41	*cis*-**T2**	−117.98	–

9.4.2 Error estimation in DCL NMR experiments

A signal-to-noise measurement was undertaken on several representative spectra from the NMR experiments undertaken on the DCLs (in **Chapter 4**) using the *.sino* function in TopSpin (Bruker BioSpin, Version 3.5 pl5, 2016). The signal region in the $^{19}F\{^{1}H\}$ NMR spectra was selected to contain the resonance associated with the 1-bromo-2-fluoro-4-nitrobenzene internal standard. The function calculated the signal-to-noise ratio (S/N) for this peak, which was used as a reference (of known concentration) for determining the concentration of all the other species in the reaction mixture. The S/N ratio was typically calculated to be around 900 and the concentration of the internal standard in the kinetic experiments was generally set at 7.8 mM, making the lowest detectable concentration between 9 and 17 µM (an error in concentration determination of $\pm$ 0.035 mM).

9.4.3 Preparation of the CD_2Cl_2 saturated with *p*TSA

An excess of *p*TSA was added to a 10 g bottle of CD_2Cl_2 and the mixture was stirred vigorously under inert argon atmosphere for 30 minutes. The excess residual solid was

filtered off under gravity and this solution was used immediately. In order to determine the concentration of the *p*TSA at saturation, the freshly prepared reaction solvent was added to a known amount of an internal standard, 1-bromo-2-fluoro-4-nitrobenzene and the mixture was analysed by quantitative 499.9 MHz ^{1}H NMR spectroscopy. The concentration estimated by integration of the peaks arising from the *p*TSA relative to the aromatic peaks on the internal standard as *ca.* 1.5 mM.

9.4.4 Kinetic simulations

A network of two replicators, **R1** and **R2**, was examined in **Chapter 4** using kinetic simulations both in the presence of dynamically-exchanging library of components and in its absence—in environment employing preformed nitrone. The kinetic simulations explored nine different conditions of $EM_{kinetic}$ and EM_{thermo} and a range of concentration conditions. The effect of varying the EM parameters and the concentration on the ratio of **R2/R1** replicators formed in the system, as well as on the conversion to all cycloadducts and the recognition-mediated products has been reported for the dynamic system in **Chapter 4**. Only the effect of these parameters on the ratio of replicators have been discussed for the reaction network in the absence of dynamic library. Therefore, the effect of the initial concentration as well as the EM_{thermo} and $EM_{kinetic}$ on the overall conversion to all cycloadducts, and the % of the **R1** and **R2** *trans* recognition-enabled products within the product pool is reported in **Figure 9.6**), for the system driven by kinetic selection purely.

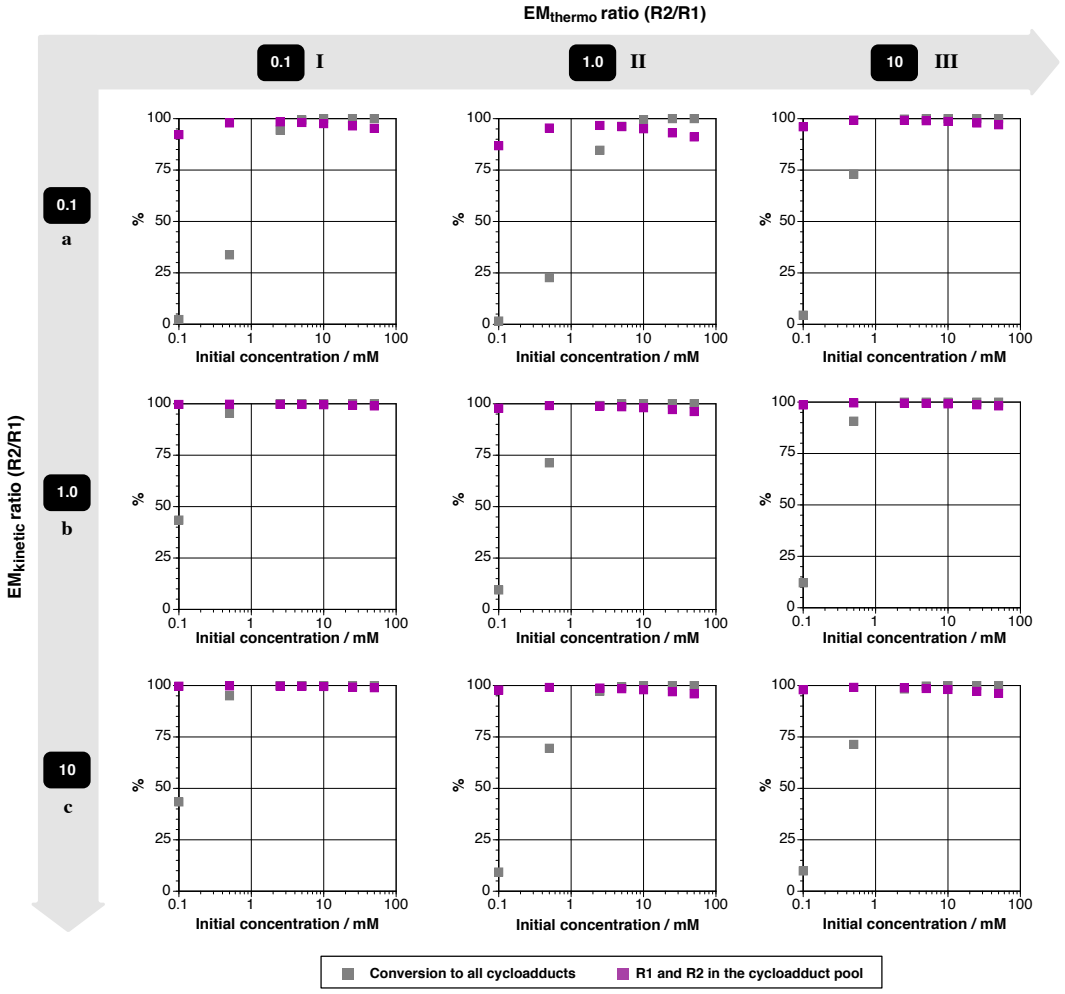


Figure 9.6 Outcome of kinetic simulations probing the influence of initial concentration conditions and the relative template duplex stability (EM_{thermo}) on the conversion to all cycloadducts (grey squares) and the % of the recognition-enabled **R1** and **R2** species in the product pool (purple squares) in reaction network, in the absence of a dynamic exchange pool, after two days. Simulations were performed using the ISOSIM mode of the SimFit software package. Note that the *x*-axis is presented in logarithmic scale in all cases.

9.5 Propagating reaction-diffusion fronts

The propagating chemical wave experiments, described in **Chapter 5**, were performed in 50 µL Hamilton 1700 series, gastight glass syringes (internal diameter = 1.03 mm), in a cold room regulated at 5 °C.

9.5.1 Determination of diffusion coefficients

As a first step in the determination of the diffusion coefficient, the diffusion delay Δ (p30) and the diffusion gradient length δ (d20) parameters were optimised for each selected compound using ^{1}H NMR experiemnts (499.9 MHz) in $CDCl_3$ at RT (*ca.* 22 °C) on a Bruker Avance III-HD 500 (pulse sequence: ledbpgp2s1d). The optimised

parameters (**Table 9.2**) were used to acquire pseudo 2D spectra using a stimulated echo pulse sequence with bipolar gradients (pulse sequence: ledbpgp2s). Each 2D DOSY dataset was opened in MestReNova (Version 10.0.2-15465, Mestrelab Research S.L., 2015). In the next step, the active spectrum mode was selected and manual phase correction applied, followed by an automatic baseline correction (full auto Bernstein polynomial). Once processes, the data was processes using Bayesian DOSY transform advanced function. The diffusion coefficient (*D*) was measured using the 'crosshair' functionality.

Table 9.2 Overview of optimised Δ (p30) and δ (d20) parameters obtained for selected maleimides, nitrone and cycloadducts using ^{1}H NMR experiemnts (499.9 MHz) spectroscopy in $CDCl_3$ at RT (*ca.* 22 °C).

	p30 / s	d20 / µs
N^{A}	0.1	800
M1	0.1	600
M2	0.1	650
T1	0.1	900
T2	0.1	950
T1^{A}	0.1	900
T2^{A}	0.1	950
T1^{A10F}	0.1	900
T2^{A10F}	0.1	950

9.5.2 Single replicator reaction-diffusion fronts

Stock solution containing the required maleimide (**M1** or **M2**) and nitrone **N^{E}** (all components at 5 mM) was prepared in fresh $CDCl_3$, using a Sartorial BP211D balance ($\pm$ 0.01 mg) and was equilibrated in the cold room (5 °C) prior to the experiment start. 50 µL of the prepared stock solution was carefully drawn into a syringe and the contents adjusted to the required volume by ejection of 5 µL of this reagent solution. In the next step, the required amount of preformed template solution (5 µL) was taken up and used to seed the propagating wave. The syringes were placed under a UV lamp (366 nm) for visualisation. Photographs of the syringes were taken with a Pentax Optio W80 12.1 megapixel digital camera, at 90 seconds intervals, for a period of four to eight hours. The photographs were processed using a freely available picture editing software Picasa 3.

9.5.3 Competition reaction-diffusion fronts employing N^{E}

Four syringes, each containing 45 µL of **N^{E}**, **M1** and **M2** in $CDCl_3$, at 5 mmol), were prepared. Three were seeded with 5 µL of preformed template: **T1** at 20 mmol, **T2** at

20 mmol, or both templates at 1 mmol each. In order to minimise the UV-catalysed hydrolysis and decomposition of the fluorescent nitrone, no photographs of chemical waves were taken if a sample was intended for NMR spectroscopic analysis. This allowed the wave to propagate for an extended period of four to seven hours, at which point the length of the observed wave was dispensed into a micro NMR tube (Bruker MatchTM 1.7 mm micro NMR tube), made up to the required volume with fresh $CDCl_3$, and analysed by 499.9 MHz ^{1}H NMR spectroscopy as soon as possible (in the meantime, the samples were kept on ice). In experiments employing DMSO-d_6, the dispensed wave sample was made up to the correct volume by addition of DMSO-d_6, instead. Double solvent suppression method was employed for ^{1}H NMR spectroscopic analysis of these samples, because of the large intensity of the residual DMSO-d_6 and water signals.

9.5.4 Control NMR experiments in DMSO-d_6

In the examination of competition networks under reaction-diffusion conditions in **Chapter 5**, DMSO-d_6 was often added to disrupt the recognition-mediated processes in the system. In order to estimate how the presence of DMSO-d_6 affects the 1,3-dipolar cycloaddition reaction between a nitrone and a maleimide, several control NMR experiments were undertaken. Specifically, these control experiments examined the reaction of nitrone $\mathbf{N^F}$ and (i) maleimide **M1** and (ii) maleimide **M2** in DMSO-$_6$ at 5 mM, at RT (23 °C). The outcome of these two control NMR experiments after *ca.* four and 16 hours are shown in **Table 9.3a** for the reaction of $\mathbf{N^F}$ with **M1** and in **Table 9.3b** for the reaction of $\mathbf{N^F}$ with **M2**.

Table 9.3 Outcome of NMR experiments examining the reaction of **(a)** $\mathbf{N^F}$ with **M1** and **(b)** $\mathbf{N^F}$ with **M2** in DMSO-d_6, as determined by 376.5 MHz ^{19}F$\{^1$H$\}$ NMR spectroscopy. Reaction conditions: 5 mM, 23 °C.

(a)

	Concentration / mM					
Time / h	**N^F**	***cis*-T1**	***trans*-T1**	Side products	[***trans*-T1**]/[***cis*-T1**]	% Conversion to ***trans*-T1** and ***cis*-T1**
4	4.68	0.12	0.22	0	1.8	6.8
18	2.92	0.39	0.82	0.74	2.1	24.2

(b)

	Concentration / mM					
Time / h	**N^F**	***cis*-T2**	***trans*-T2**	Side products	[***trans*-T2**]/[***cis*-T2**]	% Conversion to ***trans*-T2** and ***cis*-T2**
4	4.46	0.24	0.41	0	1.7	13
18	2.34	0.77	1.23	0.61	1.6	40

9.6 Rotaxane study

9.6.1 MALDI-TOF analysis of [2]rotaxanes

The **GM-R** (**Figure 9.7**) and **PM-R** (**Figure 9.8**) rotaxanes synthesised in **Chapter 6** were characterised by MALDI-TOF mass spectrometry. MALDI-TOF mass spectrometry was performed using 0.5 µL of matrix (α-cyano-4-hydroxycinnamic acid) and sample (at a 1 : 1 ratio), spotted on a plate and analysed by an ABSciex 4800 MALDI TOF/TOF (AB Sciex, Foster City, CA) mass spectrometer run in reflectron mode, equipped with a Nd:YAG 355 nm laser and calibrated using a mixture of peptides. The samples were analysed in positive MS mode with a scan breadth of 500–3000 *m/z*.

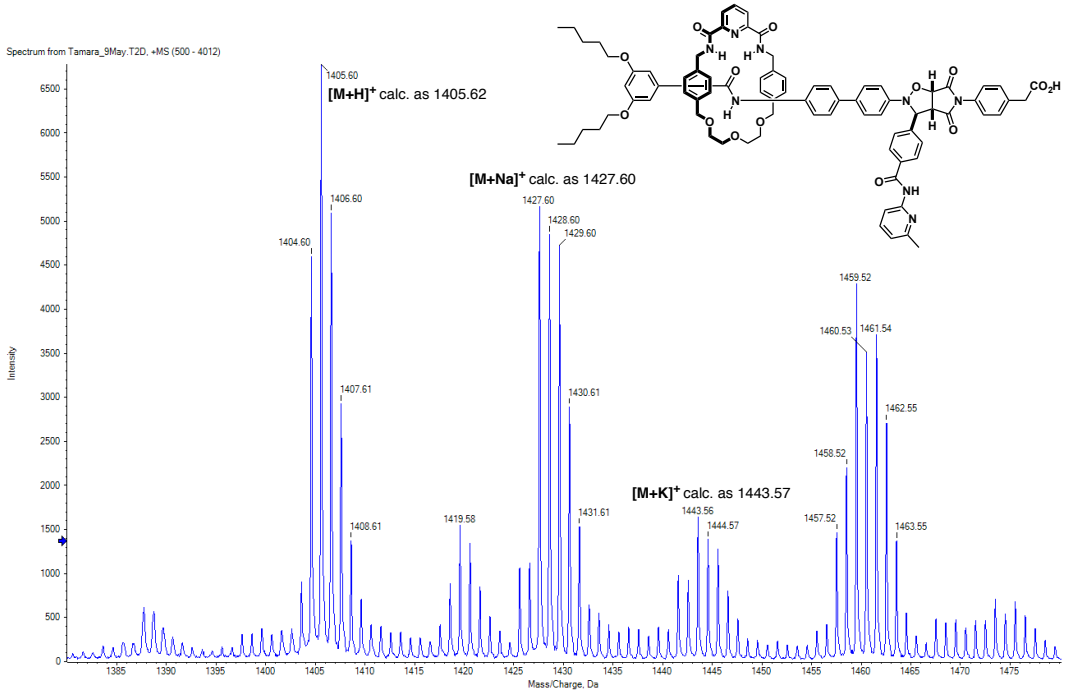


Figure 9.7 MALDI-TOF spectrum of GM [2]rotaxane.

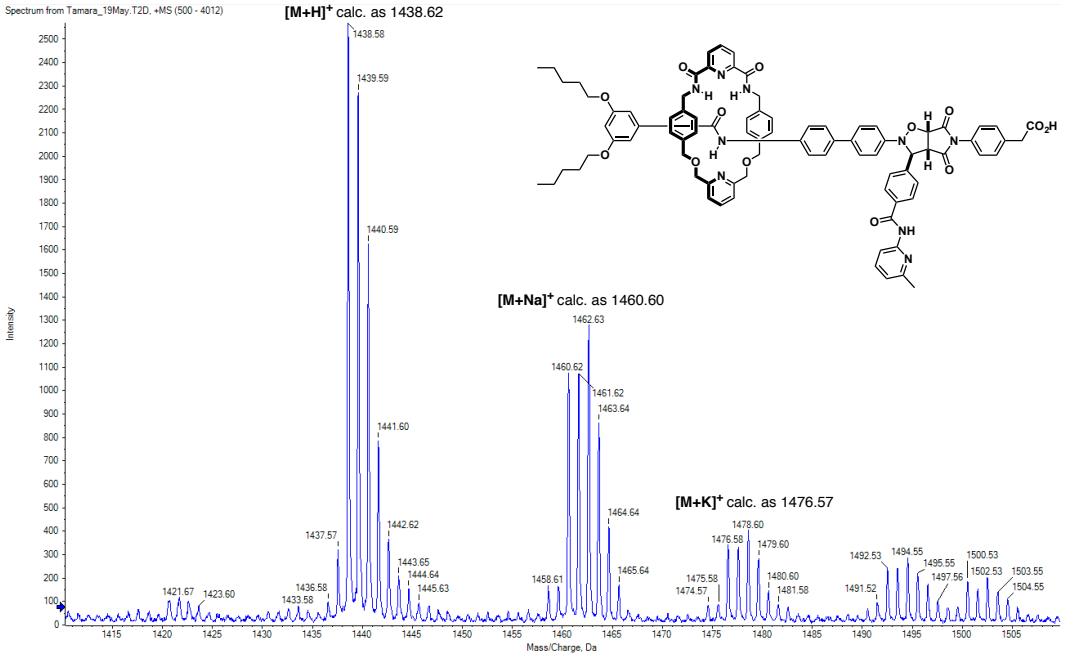


Figure 9.8 MALDI-TOF spectrum of PM [2]rotaxane.

9.6.2 Error estimation in kinetic experiments

A signal-to-noise measurement was undertaken on several spectra from the kinetic experiments undertaken on the rotaxane:thread network described in **Chapter 6**, using the *.sino* function in TopSpin (Bruker BioSpin, Version 3.2 pl3, 2013). The signal region in the ^{1}H NMR spectrum was specifically selected to contain the methyl group of the 2,4-dinitrotoluene internal standard. The function calculated the signal-to-noise ratio (S/N) for this peak, which was used as a reference (of known concentration) for determining the concentration of all the other species in the reaction mixture. The S/N ratio was typically calculated to be between 500 and 600 (corrected for 1 proton). The internal standard concentration was typically within the range of 4.5 to 7 mM, which means that the lowest concentration detectable during the kinetic experiments is from 7.5 to 14 μM, which results in error in concentration determination of (on average) ± 0.03 mM. This error also applies to the determined rate *vs* time data as the concentrations determined by NMR were directly employed in the determination of rate data.

9.6.3 Rotaxane crystallographic data

Suitable single-crystals were grown as colourless needles (**GM-R**), or plates (**PM-R**), by the slow evaporation of either MeCN/$CDCl_3$ (**GM-R**), or MeCN (**PM-R**). X-ray diffraction data for **GM-R** and **PM-R** were collected at 173 K by using a Rigaku MM-007HF High brilliance RA generator/confocal optics and Rigaku XtaLAB P100 system,

with Cu Kα radiation (λ = 1.541 87 Å). Intensity data were collected using ω and ϕ steps accumulating area detector images spanning at least a hemisphere of reciprocal space. All data were corrected for Lorentz polarization effects. A multiscan absorption correction was applied by using CrystalClear.[325,326] Structures were solved by or direct methods (SHELXD[327] or SIR2011[328]) and refined by full-matrix least-squares against F[327] (SHELXL-2013[329]). Non-hydrogen atoms were refined anisotropically, and hydrogen atoms were refined using a riding model. All calculations were performed using the CrystalStructure[330] interface. Crystals of both compounds diffracted weakly, particularly at higher angles, giving rise to poor-quality data. A number of restraints were required to obtain chemically-reasonable bond distances and angles, particularly at the periphery of the molecule. Nevertheless, an unambiguous determination of both the structures was possible. Crystallographic data for the two structures are listed in **Table 9.4**.

Table 9.4 Crystallographic data for **GM-R** and **PM-R**.

	GM-R	**PM-R**
Formula	$C_{86}H_{90}N_{10}O_{14}$	$C_{89}H_{89}N_{11}O_{13}$
Molecular weight	1487.72	1520.75
Crystal system	Monoclinic	Monoclinic
Space group	$P2_1/c$	$P2_1/c$
a (Å)	9.949(2)	9.914(5)
b (Å)	11.495(2)	11.770(5)
c (Å)	68.451(14)	73.29(3)
β (°)	93.205(4)	92.402(11)
Volume (Å^3)	7816(3)	8545(7)
Z	4	4
ρ_{calcd} (g cm^{-3})	1.264	1.182
μ (mm^{-1})	0.706	0.652
Reflections collected	71106	47075
Unique reflections (R_{int})	14168 (0.2671)	13722 (0.4063)
R_1 [I>2σ(I)]	0.1233	0.2461
wR_2 (all data)	0.4261	0.5897
GoF	1.065	1.218

9.7 Crystallographic data for T1^{A}, T1^{AF} and macrocycle 168

Suitable single-crystals were grown as colourless prisms (**T1^{A}** and **T1^{AF}**), or plates (macrocycle **168**), by the slow diffusion of Et_2O into either $CDCl_3$/PhMe (**T1^{A}**) or $CHCl_3$/THF/PhMe (**T1^{AF}**), or by the slow evaporation of MeCN/CH_2Cl_2 (macrocycle **168**). X-ray diffraction data for **T1^{A}** and **168** were collected at 173 K by using a Rigaku MM-007HF High brilliance RA generator/confocal optics and Rigaku XtaLAB P100

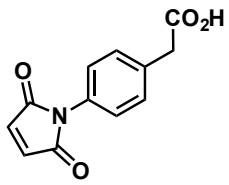
system, with Cu Kα radiation (λ = 1.541 87 Å). X-ray diffraction data for **T1^{AF}** were collected at 173 K by using a Rigaku FR-X Ultrahigh brilliance Microfocus RA generator/confocal optics and Rigaku XtaLAB P200 system, with Mo Kα radiation (λ = 0.710 75 Å). Intensity data were collected using either ω and ϕ steps (**T1^{A}** and macrocycle **168**) or just ω steps (**T1^{AF}**) accumulating area detector images spanning at least a hemisphere of reciprocal space. All data were corrected for Lorentz polarization effects. A multiscan absorption correction was applied by using CrystalClear.[325,326] Structures were solved by or direct (SIR04[331] or SIR2011[328]) or charge-flipping (Superflip[332]) methods and refined by full-matrix least-squares against F[331] (SHELXL-2013[329]). Non-hydrogen atoms were refined anisotropically, and carbon-bound hydrogen atoms were refined using a riding model. Nitrogen-bound hydrogens were located from the difference Fourier map and refined subject to distance restraint, except for the hydrogen on N14 in macrocycle **168**, which was refined using a riding model. All calculations were performed using the CrystalStructure[330] interface. Crystals of macrocycle **168** contained poorly ordered solvent molecules that could not be identified. As a result, these were modelled as eight complete and two half-occupancy water molecules. No hydrogen could be located for these water molecules. Crystallographic data for the three structures, **T1^{A}** and **T1^{AF}** and macrocycle **168**, are listed in **Table 9.5**.

Table 9.5 Crystallographic data for templates **T1^{A}** and **T1^{AF}** and macrocycle **168**.

	T1^{A}	**T1^{AF}**	**Macrocycle 168**
Formula	$C_{47}H_{35}Cl_3N_4O_6$	$C_{46}H_{33}FN_4O_6$	$C_{262}H_{288.5}N_{23.5}O_{39}$
Molecular weight	858.18	756.79	4390.80
Crystal system	Monoclinic	Monoclinic	Monoclinic
Space group	$P2_1/c$	$P2_1/c$	$P2_1/n$
a (Å)	14.5673(13)	14.579(6)	32.901(6)
b (Å)	21.0556(16)	20.986(7)	20.987(4)
c (Å)	14.7834(13)	14.686(5)	34.166(7)
β (°)	114.6320(15)	115.328(7)	94.379(3)
Volume (Å^3)	4121.8(6)	4061(3)	23523(8)
Z	4	4	4
ρ_{calcd} (g cm^{-3})	1.383	1.238	1.240
μ (mm^{-1})	2.474	0.086	0.676
Reflections collected	42832	48898	298255
Unique reflections (R_{int})	7484 (0.0456)	7439 (0.1287)	42194 (0.1973)
R_1 [I>2σ(I)]	0.0969	0.1255	0.1368
wR_2 (all data)	0.3059	0.3946	0.4076
GoF	1.095	1.049	1.057

9.8 Compound characterisation

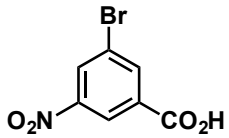
2-(4-(2,5-Dioxo-2,5-dihydro-1H-pyrrol-1-yl)phenyl)acetic acid[Ref. 127], M1



4-Aminophenylacetic acid (5.00 g, 33.1 mmol) was suspended with maleic anhydride (3.24 g, 33.0 mmol) in acetic acid (150 mL) and the reaction was stirred at RT for four hours, followed by reflux at 120 °C for three hours and standing at RT overnight. The acetic acid was removed *in vacuo* and the crude product was purified *via* silica gel flash chromatography, eluting with CH_2Cl_2:AcOH (19 : 1). The fractions containing product were combined and the solvent was removed under reduced pressure. The pure product was dissolved in minimum CH_2Cl_2, washed with Na_2CO_3 followed by HCl (1 M), the combined organic fractions were dried over $MgSO_4$, filtered and the solvent was removed *in vacuo* to yield the maleimide **M1** as a pale yellow solid (2.30 g, 30%).

M.p.: 151–162 °C (lit.[127] 151.7–153.1 °C). **^{1}H NMR** (400.1 MHz, $CDCl_3$): 7.41 (d, J = 8.4 Hz, 2H), 7.33 (d, J = 8.4 Hz, 2H), 6.86 (s, 2H), 3.69 (s, 2H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 175.8, 169.6, 134.4, 133.2, 130.6, 130.3, 126.3, 40.5. **HRMS** (ESI+) calculated for *m/z* $[M+Na]^+$ $C_{12}H_7NO_4Na$, 254.0429; found 254.0435.

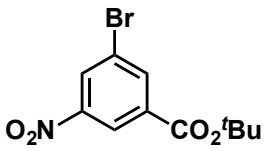
3-Bromo-5-nitrobenzoic acid[Ref. 333], 83



3-Nitrobenzoic acid (11.0 g, 65.8 mmol) was mixed with concentrated H_2SO_4 (40.0 mL) and heated to 60 °C. *N*-Bromosuccinimide (14.1 g, 79.2 mmol) was added in three portions over 30 minutes. The reaction was stirred for 1.5 hours, at which point TLC showed complete consumption of the starting material (hexane:EtOAc, 5 : 1). The reaction mixture was poured onto ice water (200 mL) and left overnight to precipitate. The product was isolated by filtration, washed with hexane and dried under high vacuum overnight to yield **83** as a white powder (15.8 g, 96%).

M.p.: 149.3–150.7 °C (lit.[333] 161.0–162.0 °C). **^{1}H NMR** (400.1 MHz, DMSO-d_6): δ 8.62–8.61 (m, 1H), 8.55 (dd, $J = 2.2$, 1.4 Hz, 1H), 8.41 (dd, $J = 1.9$, 1.4 Hz, 1H). **^{13}C NMR** (100.6 MHz, DMSO-d_6): δ 164.4, 148.7, 137.6, 134.2, 130.0, 122.8, 122.4.

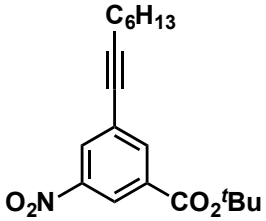
***Tert*-butyl 3-bromo-5-nitrobenzoate[Ref. 334], 84**



3-Bromo-5-nitrobenzoic acid **83** (7.38 g, 30.0 mmol) was dissolved in THF (60.0 mL) and stirred at RT under nitrogen. Di-*t*-Bu dicarbonate (19.6 g, 89.8 mmol) and dimethylaminopyridine (840 mg, 6.88 mmol) were added and the reaction stirred under inert atmosphere for 16 hours before being poured onto ice water (200 mL). The solution was made basic (pH 10) by addition of Na_2CO_3 and the product was extracted into CH_2Cl_2 (3×100 mL). The combined organic fractions were dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude product was purified by silica gel flash chromatography with cyclohexane:EtOAc (10 : 1) to yield the pure **84** as a white crystalline solid (7.20 g, 79%).

M.p.: 76.4–77.5 °C. **^{1}H NMR** (499.9 MHz, $CDCl_3$): δ 8.71 (dd, $J = 2.1$, 1.4 Hz, 1H), 8.51 (t, $J = 2.0$ Hz, 1H), 8.41 (t, $J = 1.6$ Hz, 1H), 1.62 (s, 9H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 162.4, 148.8, 138.3, 135.3, 130.1, 123.2, 123.0, 83.5, 28.2.

***Tert*-butyl 3-nitro-5-(oct-1-yn-1-yl)benzoate, 85**

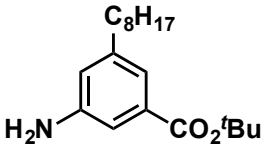


Tert-butyl 3-bromo-5-nitrobenzoate **84** (4.00 g, 12.1 mmol) was dissolved in vigorously degassed Et_3N (120 mL) under nitrogen atmosphere. Octyne (5.74 g, 52.1 mmol), $PdCl_2(PPh_3)_2$ (0.600 g, 0.850 mmol), copper iodide (0.170 g, 0.920 mmol) and PPh_3 (0.500 g, 1.85 mmol) were added and the solution heated to 80 °C for 44 hours. The reaction was allowed to cool to RT, filtered through celite, and the solvent removed *in vacuo*. The crude material was purified by silica gel flash chromatography, using

gradient elution: from cyclohexane to cyclohexane:EtOAc (15 : 1) to yield **85** as light brown oil (3.80 g, 86%).

^{1}H NMR (400.1 MHz MHz, $CDCl_3$): δ 8.64 (dd, J = 2.2, 1.5 Hz, 1H), 8.33 (dd, J = 2.2, 1.6 Hz, 1H), 8.25 (t, J = 1.5 Hz, 1H), 2.43 (t, J = 7.1 Hz, 2H), 1.61 (s, 9H), 1.49–1.41 (m, 2H), 1.37–1.27 (m, 6H), 0.91 (t, J = 7.1 Hz, 3H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 163.2, 148.2, 138.0, 133.9, 129.6, 126.4, 123.1, 94.7, 82.9, 78.1, 31.5, 28.8, 28.5, 28.2, 22.7, 19.5, 14.2.

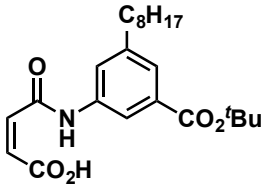
***Tert*-butyl 3-amino-5-octylbenzoate, 86**



Tert-butyl 3-nitro-5-(oct-1-yn-1-yl)benzoate **85** (0.500 g, 1.51 mmol) was dissolved in HPLC grade methanol (22.5 mL) and THF (2.50 mL), the solution was sonicated and run through an H-cube continuous flow hydrogenation reactor (1 mL/min, 50 bar, 50 °C) loaded with Pd/C 10% catalyst cartridge. The conversion to product was monitored by TLC (PhMe/MeOH/Et_3N, 10 : 1 : 0.1). Upon complete conversion to product (generally three runs were required), the solution was reduced *in vacuo*, affording **86** as a yellow oil, sufficiently pure for further conversion (0.41 g, 90%).

^{1}H NMR (300.1 MHz, $CDCl_3$): δ 7.24–7.23 (m, 1H), 7.15 (t, J = 1.8 Hz, 1H), 6.70 (t, J = 1.8 Hz, 1H), 3.80 (s, 2H), 2.54 (t, J = 7.8 Hz, 2H), 1.63–1.53 (m, 11H), 1.36–1.20 (m, 10H), 0.88 (t, J = 6.7 Hz, 3H). **^{13}C NMR** (75.5 MHz; $CDCl_3$): δ 166.2, 145.6, 144.4, 133.1, 120.5, 119.7, 113.8, 80.9, 36.0, 32.0, 31.4, 29.6, 29.5, 29.4, 28.3, 22.8, 14.3. **HRMS** (CI+) *m/z* calculated for $C_{19}H_{32}NO_2$ $[M+H]^+$ 306.2433, found 306.2438.

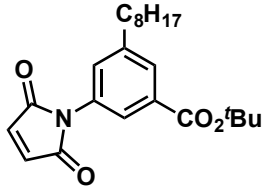
(*Z*)-4-((3-(*tert*-butoxycarbonyl)-5-octylphenyl)amino)-4-oxobut-2-enoic acid, 87



Tert-butyl 3-amino-5-octylbenzoate **86** (1.10 g, 3.60 mmol) was dissolved in dry THF (25.0 mL) and maleic anhydride (0.380 g, 3.60 mmol) was added and the mixture stirred under nitrogen atmosphere for six hours at RT, at which point TLC analysis (hexane:EtOAc, 3 : 1) showed complete conversion of starting materials to product. The solution was concentrated *in vacuo* to yield the intermediate uncyclised product **87** as a pale yellow solid, sufficiently pure for immediate further reaction (quantitative conversion).

^{1}H NMR (400.1 MHz MHz, $CDCl_3$): δ 9.57 (s, 1H), 7.96 (t, J = 1.8 Hz, 1H), 7.77 (t, J = 1.8 Hz, 1H), 7.66 (t, J = 1.5 Hz, 1H), 6.67 (d, J = 12.9 Hz, 1H), 6.45 (d, J = 12.8 Hz, 1H), 2.64 (t, J = 7.8 Hz, 2H), 1.64–1.56 (m, 10H), 1.34–1.21 (m, 10H), 0.86 (t, J = 6.9 Hz, 3H).

***Tert*-butyl 3-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)-5-octylbenzoate, 88**

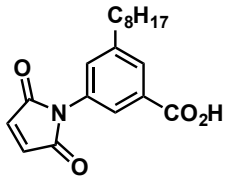


Compound **87** (3.60 mmol) and sodium acetate (0.880 g, 10.7 mmol) were dissolved in acetic anhydride (20.0 mL) and PhMe (10.0 mL) and the mixture was refluxed until TLC analysis confirmed complete conversion to product (2 hours, hexane:EtOAc, 3 : 2). The reaction was cooled to RT, remaining undissolved sodium acetate was filtered off and acetic anhydride and PhMe were removed *in vacuo*. H_2O (20.0 mL) was added to the mixture and the product extracted with CH_2Cl_2, the combined organic fractions washed with K_2CO_3 solution, dried over $MgSO_4$, and the solvent removed under reduced pressure to afford **88** as a brown oil (1.20 g, 86%).

^{1}H NMR (300.1 MHz, $CDCl_3$): δ 7.82 (t, J = 1.6 Hz, 1H), 7.74 (t, J = 1.8 Hz, 1H), 7.29 (t, J = 1.8 Hz, 1H), 6.87 (s, 2H), 2.67 (t, J = 7.8 Hz, 2H), 1.66–1.60 (m, 2H), 1.58 (s, 9H), 1.36–1.22 (m, 10H), 0.89–0.85 (m, 3H). **^{13}C NMR** (100.6 MHz, $CDCl_3$):

δ 169.5, 165.1, 144.5, 134.4, 133.2, 131.2, 130.3, 129.2, 124.8, 81.6, 35.8, 32.0, 31.3, 29.5, 29.4, 28.3, 28.3, 22.8, 14.3. **HRMS** (ESI+) *m/z* calculated for $[M+Na]^+$ $C_{23}H_{31}NO_4Na$ 408.2151, found 408.2149.

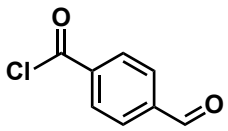
3-(2,5-Dioxo-2,5-dihydro-1H-pyrrol-1-yl)-5-octylbenzoic acid, M2



The protected compound **88** (1.20 g, 3.11 mmol) was dissolved in dry CH_2Cl_2 (15.0 mL) and stirred under nitrogen. TFA (4.00 mL) was added and the reaction stirred at RT. After two hours, TLC analysis has shown that the reaction had gone to completion and was quenched by addition of H_2O (100 mL) and the product was extracted into CH_2Cl_2 and the combined organic fractions were dried over $MgSO_4$, filtered and the solvent removed under reduced pressure to give the crude product that was purified by silica gel flash chromatography (hexane:EtOAc, 3 : 2) to yield the final product **M2** as a beige solid (0.500 g, 49%).

M.p.: 128.6–130.0 °C. **^{1}H NMR** (400.1 MHz MHz, $CDCl_3$): δ 7.94–7.92 (m, 2H), 7.42 (t, J = 1.8 Hz, 1H), 6.89 (s, 2H), 2.71 (t, J = 7.8 Hz, 2H), 1.70–1.62 (m, 2H), 1.39–1.24 (m, 10H), 0.91–0.85 (m, 3H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 170.9, 169.4, 144.9, 134.5, 131.59, 131.45, 130.3, 129.8, 125.3, 35.8, 32.0, 31.3, 29.5, 29.4, 29.4, 22.8, 14.3.

4-Formylbenzoyl chloride[Ref. 181], 89

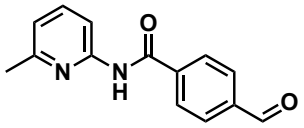


4-Formylbenzaldehyde (3.98 g, 16.6 mmol) was suspended in dry PhMe (10.0 mL) under inert atmosphere. $SOCl_2$ (50.0 mL) was slowly injected at ambient temperature and the mixture was refluxed at 110 °C for 24 hours or until TLC analysis showed complete consumption of the starting material. The obtained brown solution was cooled to RT and the solvent was removed under reduced pressure (removal of excess $SOCl_2$ was facilitated by repeated co-evaporation with PhMe) to yield the crude acid chloride

89 as a brown solid, sufficiently pure for immediate further use (quantitative conversion).

^{1}H NMR (400.1 MHz MHz, $CDCl_3$): δ 10.15 (s, 1H), 8.29 (d, $J = 8.5$ Hz, 2H), 8.03 (d, $J = 8.5$ Hz, 2H).

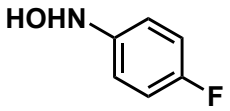
4-Formyl-*N*-(6-methylpyridin-2-yl)benzamide[Ref. 127], A



2-Amino-6-picoline (3.86 g, 35.6 mmol) and Et_3N (10.0 mL) were dissolved in dry CH_2Cl_2 (30.0 mL) and cooled on ice under nitrogen atmosphere. 4-Formylbenzoyl chloride **89** (5.00 g, 29.7 mmol) was dissolved in dry CH_2Cl_2 (60.0 mL) and was added dropwise to the cooled solution (over 2 hour period). Once the addition was complete, the solution was allowed to warm to RT and was left to stir overnight. After 18 hours, TLC analysis showed complete consumption of the starting material. The excess amines were removed by washing with HCl (1 M, 2 $\times$ 50.0 mL). The aqueous fraction was repeatedly washed with CH_2Cl_2 and the resulting combined organic fractions were washed with $NaHCO_3$ and brine, dried with $MgSO_4$, the solution filtered and solvent removed under reduced pressure. The crude aldehyde was purified via silica gel flash chromatography (gradient elution with CH_2Cl_2:EtOAc, from 6 : 1 to 3 : 1) to yield the final aldehyde **A** as a white solid (3.80 g, 53%).

M.p.: 106.7–108.0 °C (lit.[127] 108.4–109.6 °C). **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 10.06 (s, 1H), 9.03 (s, br, 1H), 8.15 (d, $J = 8.2$ Hz, 1H), 8.05–8.02 (m, 2H), 7.95–7.92 (m, 2H), 7.63 (t, $J = 7.9$ Hz, 1H), 6.91 (d, $J = 7.4$ Hz, 1H), 2.35 (s, 3H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 191.5, 164.6, 157.3, 150.5, 139.6, 139.0, 138.7, 130.1, 128.0, 120.1, 111.2, 24.2. **HRMS** (ESI+) *m/z* calculated for $[M+H]^+$ $C_{14}H_{13}N_2O_2$, 241.0972; found 241.0977.

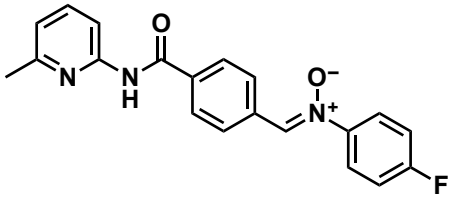
***N*-(4-Fluorophenyl)hydroxylamine[Ref. 126, 181], Z**



Fluoro-4-nitrobenzene (1.00 g, 7.86 mmol) was dissolved in dry THF (50.0 mL) and rhodium catalyst (0.200 g, 5% wt. on carbon) was added. Hydrazine monohydrate (0.450 mL, 0.464 mg, 9.28 mmol) was added and the reaction followed by TLC (hexane:EtOAc, 3 : 1). After completion (Around 40 minutes), $MgSO_4$ was added and the reaction stirred for further five minutes. The resulting solution was filtered through celite, washed with dry THF and concentrated *in vacuo* at 25 °C. Recrystallisation from EtOAc/hexane afforded the hydroxylamine product **Z** as a white crystalline solid (0.790 g, 87%).

^{1}H NMR (500.1 MHz, $CDCl_3$): δ 7.00–6.95 (m, 4H), 6.74 (br s, 1H), 5.97 (br s, 1H). **^{13}C NMR** (75.5 MHz, $CDCl_3$): δ 158.7 (d, J = 240.2 Hz), 145.5 (d, J = 2.3 Hz), 116.6 (d, J = 7.9 Hz), 115.7 (d, J = 22.7 Hz). **^{19}F$\{^1$H$\}$ NMR** (376.5 MHz, $CDCl_3$) δ −121.6.

(*Z*)-*N*-(4-Fluorophenyl)-1-(4-((6-methylpyridin-2-yl)carbamoyl)phenyl)methanimine oxide[Ref. 126], nitrone N^F or AZ

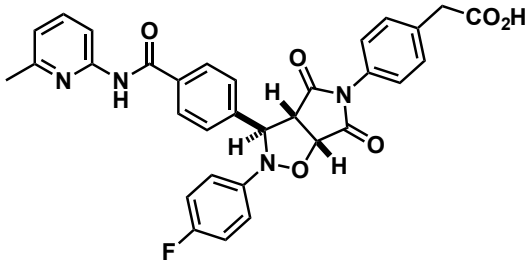


N-(4-Fluorophenyl)hydroxylamine **Z** (150 mg, 1.18 mmol) and aldehyde **A** (290 mg, 1.21 mmol) were dissolved in EtOH (5.00 mL) and the mixture was sonicated until the solids completely dissolved. The reaction was left unstirred in a freezer at −20 °C for 18 hours. The resulting precipitate was filtered and washed with cold hexane to afford the nitrone **N^F** (**AZ**) as an off-white solid (330 mg, 80%).

M.p.: 215.7–217.0 °C (lit. 217.2–218.3 °C). **^{1}H NMR** (400.3 MHz, $CDCl_3$): δ 8.55–8.48 (m, 3H), 8.19 (d, J = 8.2 Hz, 1H), 8.05–8.02 (m, 2H), 7.97 (s, 1H), 7.84–7.78 (m, 2H), 7.67 (t, J = 7.9 Hz, 1H), 7.23–7.17 (m, 2H), 6.97–6.95 (m, 1H), 2.49 (s, 3H). **^{13}C NMR** (125.7 MHz, $CDCl_3$) δ 164.7, 163.4 (d, J = 287.8 Hz),

157.2, 150.7, 145.3, 138.9, 135.9, 133.9, 133.4, 129.2, 127.7, 123.9 (d, $J = 9.1$ Hz), 119.8, 116.4 (d, $J = 23.5$ Hz), 111.1, 24.2. **^{19}F{^{1}H} NMR** (470.3 MHz, $CDCl_3$): δ −109.6. **HRMS** (ESI+) calculated for *m/z* $[M+Na]^+$ $C_{20}H_{16}N_3O_2FNa$, 372.1124; found 372.1114.

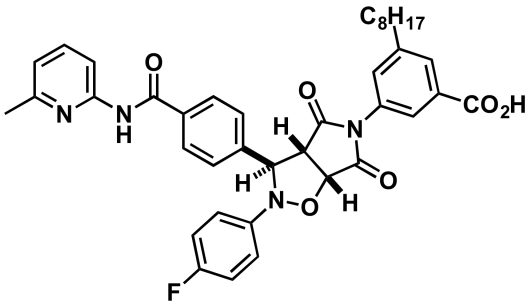
Cycloadduct *trans*-T1[Ref. 126]



Nitrone $\mathbf{N^F}$ (66.0 mg, 0.190 mmol) and maleimide **M1** (44.0 mg, 0.190 mmol) were dissolved in $CDCl_3$ (9.00 mL), the mixture sonicated and left in a fridge at 4 °C for 13 days. The resulting precipitate was filtered under suction and the crude product was recrystallised from CH_2Cl_2/hexane, affording the *trans*-**T1** product as a white solid (56.0 mg, 51%).

M.p.: 234 °C (decomp.) (lit.[126] 203.5–204.8 °C). 1HF NMR (499.9 MHz, DMSO-d_6): δ 12.44 (s, 1H), 10.75 (s, 1H), 8.07–8.05 (m, 2H), 8.01 (d, $J = 8.2$ Hz, 1H), 7.73 (t, $J = 7.9$ Hz, 1H), 7.66–7.64 (m, 2H), 7.30–7.29 (m, 2H), 7.26–7.22 (m, 2H), 7.13—7.10 (m, 2H), 7.03 (d, $J = 7.5$ Hz, 1H), 6.69–6.67 (m, 2H), 5.95 (s, 1H), 5.45 (d, $J = 7.4$ Hz, 1H), 4.17 (d, $J = 7.4$ Hz, 1H), 3.60 (s, 2H), 2.45 (s, 3H). **^{13}C NMR** (125.7 MHz, DMSO-d_6): δ 174.4, 173.3, 172.4, 165.5, 157.9 (d, $J = 235$ Hz), 156.6, 151.5, 145.1 (d, $J = 1.9$ Hz), 142.6, 138.4, 135.8, 133.6, 130.0, 128.3, 127.2, 126.0, 119.1, 116.3 (d, $J = 7.8$ Hz), 115.6 (d, $J = 22.7$ Hz), 111.7, 77.7, 68.4, 56.4, 40.1, 23.6. **^{19}F{^{1}H} NMR** (470.3 MHz, $CDCl_3$): δ −120.6. **MS** (ESI+) *m/z* for $[M+H]^+$ $C_{32}H_{26}FN_4O_6$ 581.18.

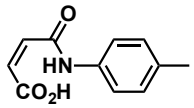
Cycloadduct *trans*-T2



Nitrone $\mathbf{N^F}$ (66.0 mg, 0.190 mmol) and maleimide **M2** (62.0 mg, 0.190 mmol) were dissolved in $CDCl_3$ (9.00 mL), the mixture sonicated and left in a fridge at 4 °C for 13 days. After this time, the solvent was evaporated under a flow of compressed air and the crude product was recrystallised from CH_2Cl_2/hexane, affording the *trans*-**T2** product as a white solid (42.0 mg, 33%).

M.p.: $>$ 232 °C (decomp.). $^{\mathbf{1}}$**H NMR** (499.9 MHz, DMSO-d_6): δ 13.20 (s, 1H), 10.76 (s, 1H), 8.08–8.07 (m, 2H), 8.02 (d, $J = 8.2$ Hz, 1H), 7.77 (s, 1H), 7.73 (t, $J = 7.9$ Hz, 1H), 7.67–7.66 (m, 2H), 7.30–7.27 (m, 3H), 7.13–7.07 (m, 2H), 7.03 (d, $J = 7.5$ Hz, 1H), 6.49 (s, 1H), 6.01 (s, 1H), 5.46 (d, $J = 7.4$ Hz, 1H), 4.16 (d, $J = 7.4$ Hz, 1H), 2.58 (t, $J = 7.6$ Hz, 2H), 2.46 (s, 3H), 1.53–1.48 (m, 2H), 1.31–1.20 (m, 10H), 0.86 (t, $J = 6.9$ Hz, 3H). $^{\mathbf{13}}$**C NMR** (125.7 MHz, DMSO-d_6): δ 174.3, 173.1, 166.4, 165.5, 158.1 (d, $J = 239$ Hz), 156.6, 151.5, 145.3 (d, $J = 2.1$ Hz), 143.9, 142.7, 138.4, 133.6, 131.7, 131.6, 130.2, 129.4, 128.4, 127.1, 124.9, 119.1, 116.2 (d, $J = 7.8$ Hz), 115.7 (d, $J = 22.4$ Hz), 111.7, 78.0, 68.4, 56.5, 34.6, 31.3, 30.8, 28.8, 28.7, 28.6, 23.6, 22.1, 14.0. $^{\mathbf{19}}$**F**$\{^{\mathbf{1}}$**H**$\}$ **NMR** (470.3 MHz, $CDCl_3$): δ -119.7. **HRMS** (ESI+) *m/z* calculated for $[M+H]^+$ $C_{39}H_{40}FN_4O_6$ 679.2932, found 679.2904.

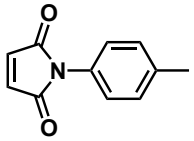
(*Z*)-4-Oxo-4-(*p*-tolylamino)but-2-enoic acid, 91



4-Aminotoluene (2.14 g, 20.0 mmol) and maleic anhydride (1.96 g, 20.0 mmol) were reacted in AcOH (50.0 mL) at RT for three hours. The yellow precipitate formed during the reaction was filtered, washed with Et_2O and dried under vacuum to yield the uncyclised product **91** in sufficient purity for direct conversion (2.52 g, 61%).

^{1}H NMR (400.1 MHz, $CDCl_3$): δ 7.77 (s, 1H), 7.47–7.44 (m, 2H), 7.23–7.19 (m, 2H), 6.47 (d, J = 12.8 Hz, 1H), 6.35 (d, J = 12.8 Hz, 1H), 2.36 (s, 3H). **^{13}C NMR** (75.5 MHz, DMSO-d_6): δ 166.7, 163.0, 135.9, 133.0, 131.6, 130.7, 129.2, 119.6, 20.5. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{11}H_{11}NO_3Na$ 228.0631, found 228.0633.

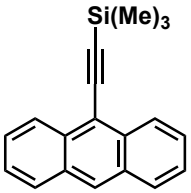
1-(*p*-Tolyl)-1*H*-pyrrole-2,5-dione[Ref. 334, 241, 335], M3



Uncyclised intermediate **91** (0.500 g, 2.44 mmol) was dissolved in acetic anhydride (15.0 mL), NaOAc (0.598 g, 7.29 mmol) was added and the reaction mixture was refluxed at 110 °C for three hours. The reaction was cooled and stirred at RT for a further hour. Residual undissolved NaOAc was filtered off under gravity and the filtrate was diluted with H_2O. The organic product was extracted into CH_2Cl_2, dried over $MgSO_4$ and concentrated under reduced pressure, to produce the desired recognition-disabled maleimide **M3** as a yellow solid (0.24 g, 52%).

M.p.: 149.3–151.6 °C. (lit.[335] 180–181 °C). **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 7.28–7.26 (m, 2H), 7.22–7.19 (m, 2H), 6.84 (s, 2H), 2.38 (s, 3H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 169.8, 138.2, 134.3, 129.9, 128.6, 126.2, 21.3. **HRMS** (CI+): *m/z* calculated for $[M+H]^+$ $C_{11}H_{10}NO_2$ 188.0706; found 188.0715.

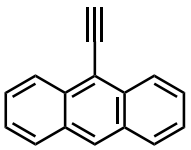
Anthracene 9-ylethynyl trimethylsilane[Ref. 336, 337], 96



9-Bromoanthracene (7.50 g, 29.2 mmol) was dissolved in Et_3N (300 mL) and the mixture was vigorously degassed for 10 minutes with argon, at RT. PPh_3 (0.760 g, 2.90 mmol), CuI (0.560 g, 2.92 mmol) and $PdCl_2(PPh_3)_2$ (0.680 g, 0.970 mmol) were added and the solution was degassed for further 30 minutes under argon. Trimethylsilylacetylene (6.25 mL, 43.6 mmol) was injected, solution degassed again for 30 more minutes and the reaction stirred for 18 hours at 80 °C, under argon atmosphere. The resulting solution was filtered through celite and concentrated *in vacuo*. The crude product was recrystallised from Et_2O/MeOH to yield pure **96** (4.38 g). The remaining crude product was purified using silica gel flash chromatography (hexane:EtOAc, 0 to 2%) to yield further 2.46 g of **96**. Overall, the TMS-protected product **96** was obtained as a bright orange solid (6.84 g, 85% over two steps).

M.p.: 80.3–83.5 °C. **^{1}H NMR** (300.1 MHz, $CDCl_3$): δ 8.54–8.51 (m, 2H), 8.42 (s, 1H), 7.98–7.95 (m, 2H), 7.58–7.44 (m, 4H), 0.42 (s, 9H). **^{13}C NMR** (75.5 MHz, $CDCl_3$): δ 132.6, 130.8, 128.3, 127.6, 126.5, 126.4, 125.5, 117.1, 105.9, 101.3, 0.3. **HRMS** (CI+) *m/z* calculated for $[M+H]^+$ $C_{19}H_{19}Si$, 275.1256; found, 275.1249.

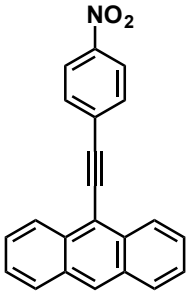
9-Ethynylanthracene, 97



(Anthracen-9-ylethynyl)trimethylsilane **96** (3.00 g, 11.5 mmol) was dissolved in MeOH (50.0 mL) and THF (50.0 mL). K_2CO_3 (7.60 g, 55.0 mmol) was added and the solution stirred for three hours at RT. H_2O (100 mL) was added to the resulting solution, and the product was extracted into CH_2Cl_2 (2 $\times$ 100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo* to 10% of the original volume to give the intermediate 9-ethynylanthracene **97** as brown-red oil that was sufficiently pure for further conversion (quantitative conversion).

^{1}H NMR (300.1 MHz, $CDCl_3$): δ 8.54–8.51 (m, 2H), 8.29 (s, 1H), 7.86–7.83 (m, 2H), 7.46–7.32 (m, 4H), 3.92 (s, 1H).

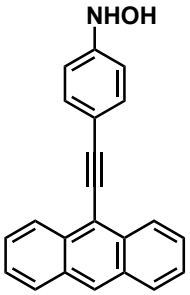
9-((4-Nitrophenyl)ethynyl)anthracene[Ref. 336, 338], 98



1-Iodo-4-nitrobenzene (2.44 g, 9.92 mmol) was dissolved in Et_3N (180 mL), and PPh_3 (0.260 g, 0.500 mmol), copper iodide (0.190 g, 0.500 mmol) and $PdCl_2(PPh_3)_2$ (0.350 g, 0.250 mmol) were added and the solution was degassed for 30 minutes under argon. Crude 9-ethynylanthracene **97** (11.5 mmol) was dissolved separately in Et_3N (20.0 mL) and degassed with argon for 20 minutes. The 9-ethynylanthracene **97** solution was added to the 1-iodo-4-nitrobenzene mixture and the reaction was stirred at (55 °C) for 66 hours. The resulting mixture was filtered through celite and concentrated *in vacuo*. The crude product was purified using silica gel flash chromatography (hexane:EtOAc, 10 : 1) to afford the compound **98** as a red crystalline solid (0.900 g, 19%).

M.p.: 203.2–205.6 °C (lit.[338] 217–219 °C). **^{1}H NMR** (300.1 MHz, $CDCl_3$): δ 8.63 (d, $J = 6.8$ Hz, 2H), 8.56 (s, 1H), 8.38–8.35 (m, 2H), 8.10 (d, $J = 6.8$ Hz, 2H), 7.96–7.93 (m, 2H), 7.71–7.58 (m, 4H). **^{13}C NMR** (100.6 MHz, $CDCl_3$) δ 147.0, 132.9, 132.2, 131.1, 130.5, 129.2, 128.9, 127.2, 126.4, 125.9, 123.9, 115.8, 98.8, 91.9.

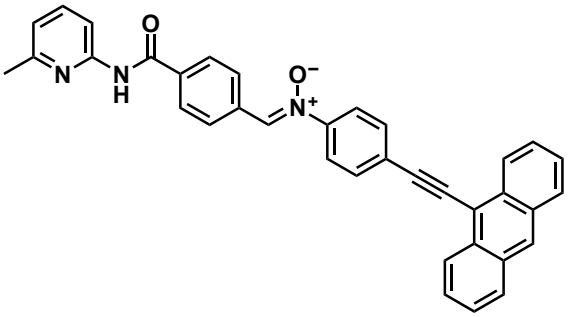
_N_-(4-(Anthracen-9-ylethynyl)phenyl)hydroxylamine[Ref. 241], 99



9-((4-Nitrophenyl)ethynyl)anthracene **98** (86.0 mg, 0.266 mmol) was dissolved in dry THF (10.0 mL) and rhodium catalyst (40.0 mg, 5% wt, on carbon) was added. After 5 minutes, hydrazine monohydrate (80.0 µL, 77.5 µg, 1.55 mmol) was added dropwise to the suspension at RT under argon atmosphere. Reaction progress was monitored by TLC (cyclohexane:EtOAc, 3 : 1) and the product spot was visualised by staining with ninhydrin solution, followed by heating. Upon completion (*ca.* 25 min), the mixture was filtered through celite (washing with dry THF) and the solvent was removed *in vacuo* at 25 °C. Crude product was dried under high vacuum overnight to give the product **99** as an orange-brown solid, sufficiently pure for immediate further reaction (quantitative conversion).

^{1}H NMR (300.1 MHz, $CDCl_3$): δ 8.65 (d, $J = 8.5$ Hz, 2H), 8.42 (s, 1H), 8.02 (d, $J = 8.5$ Hz, 2H), 7.70 (d, $J = 8.5$ Hz, 2H), 7.62 (m, 4H) 7.07 (d, $J = 8.5$ Hz, 2H), 6.98 (br s, 1H), 5.31 (br s, 1H). **MS** (ESI−) *m/z* $[M-H]^-$ 308.04.

(*Z*)-*N*-(4-(Anthracen-9-ylethynyl)phenyl)-1-(4-((6-methylpyridin-2-yl)carbamoyl)-phenyl)methanimine oxide[Ref. 241], nitrone N^E

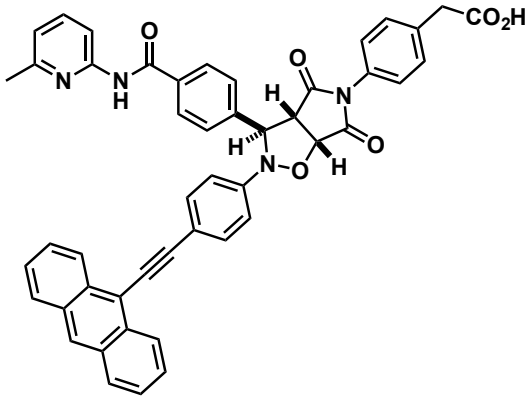


Crude *N*-(4-(anthracen-9-ylethynyl)phenyl)hydroxylamine **99** (50.0 µg, 0.161 mmol) and aldehyde **A** (34.8 µg, 0.145 mmol) were dissolved in EtOH (3.50 mL) and $CHCl_3$

(2.50 mL) and the solution was left to stand at 5 °C for three days. The resulting precipitate was filtered and washed with cold hexane to give $\mathbf{N^E}$ as an orange solid (35.6 mg, 46%).

M.p.: 201.5 °C (decomp.). **^{1}H NMR** (400.1 MHz MHz, $CDCl_3$): δ 8.63 (d, $J = 6.0$ Hz, 2H), 8.55 (d, $J = 6.6$ Hz, 2H), 8.48 (s, 1H), 8.09 (s, 1H), 8.08–8.04 (m, 5H), 7.92–7.86 (m, 4H), 7.76 (t, $J = 7.5$ Hz, 1H), 7.70–7.49 (m, 5H), 6.91 (d, $J = 7.2$ Hz, 1H), 2.53 (s, 3H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 164.4, 150.6, 148.2, 139.2, 136.5, 134.1, 133.5, 132.8, 132.5, 131.2, 129.3, 128.9, 128.5, 127.7, 127.0, 126.6, 125.9, 122.0, 119.7, 116.4, 111.2, 97.5, 91.9, 23.9. **MS** (ESI+) *m/z* $[M+Na]^+$ 554.95.

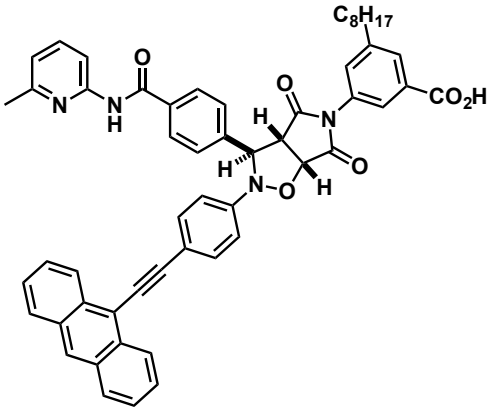
Cycloadduct *trans*-T1$^{E[Ref.\ 241]}$



Hydorxylamine **99** (50.0 mg, 0.160 mmol), aldehyde **A** (39.0 mg, 0.160 mmol) and maleimide **M1** (37.0 mg, 0.160 mmol) were weighed out into a 10.0 mL glass vial and dissolved in $CDCl_3$ (6.20 mL). The vial was covered in foil and put in the freezer overnight (−20 °C) and then in the fridge for 24 days (4 °C). The resulting precipitate was filtered, affording *trans*-$\mathbf{T1^E}$ as an orange powder (89.0 mg, 73%).

M.p.: 215 °C (decomp.). **^{1}H NMR** (499.9 MHz, DMSO-d_6): δ 12.37 (s, 1H), 10.76 (s, 1H), 8.69 (s, 1H), 8.59 (d, $J = 8.6$ Hz, 2H), 8.18–8.17 (m, 2H), 8.11–8.09 (m, 2H), 8.02 (d, $J = 8.3$ Hz, 1H), 7.78–7.68 (m, 7H), 7.63–7.60 (m, 2H), 7.40–7.38 (m, 2H), 7.29–7.28 (m, 2H), 7.03 (d, $J = 7.3$ Hz, 1H), 6.72–6.70 (m, 2H), 6.13 (s, 1H), 5.53 (d, $J = 7.4$ Hz, 1H), 4.22 (d, $J = 7.5$ Hz, 1H), 3.54 (s, 2H), 2.45 (s, 3H). **^{13}C NMR** (125.7 MHz, DMSO-d_6): δ 174.7, 173.6, 172.8, 165.9, 157.1, 152.0, 149.5, 142.9, 138.9, 136.3, 134.1, 133.0, 132.2, 131.2, 130.5, 130.4, 129.4, 128.8, 128.3, 127.8, 127.6, 126.6, 126.6, 126.5, 119.6, 116.7, 116.5, 115.2, 112.2, 101.2, 85.5, 78.3, 68.4, 56.8, 40.1, 24.1. **HRMS** (ESI+) *m/z* calculated for $[M+H]^+$ $C_{48}H_{35}N_4O_6$ 763.2557, found 763.2557.

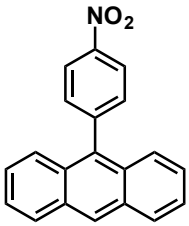
Cycloadduct *trans*-T2^{E}



Hydorxylamine **99** (50.0 mg, 0.160 mmol), aldehyde **A** (59.0 mg, 0.160 mmol) and maleimide **M2** (53.0 mg, 0.160 mmol) were weighed out into a 10.0 mL glass vial and dissolved in $CDCl_3$ (6.20 mL). The vial was covered in foil and put in the freezer overnight (−20 °C) and then in the fridge for 24 days (4 °C). The resulting precipitate was filtered, and the crude product was purified by recrystallisation from cold CH_2Cl_2/hexane, to give *trans*-**T2^{E}** as an orange powder (81.0 mg, 59%).

M.p.: 231 °C (decomp.). **^{1}H NMR** (499.9 MHz, DMSO-d_6): δ 12.94 (s, 1H), 10.79 (s, 1H), 8.67 (s, 1H), 8.62–8.59 (m, 2H), 8.18–8.12 (m, 4H), 8.04 (d, J = 8.2 Hz, 1H), 7.81–7.79 (m, 2H), 7.76–7.732 (m, 3H), 7.72–7.67 (m, 3H), 7.65–7.59 (m, 2H), 7.55 (t, J = 1.6 Hz, 1H), 7.47–7.44 (m, 2H), 7.04 (d, J = 7.5 Hz, 1H), 6.21 (s, 1H), 6.04 (t, J = 1.6 Hz, 1H), 5.51 (d, J = 7.3 Hz, 1H), 4.23 (d, J = 7.5 Hz, 1H), 2.46 (s, 3H), 1.35–1.27 (m, 2H), 0.95–0.88 (m, 2H), 0.86–0.63 (m, 10H), 0.54–0.47 (m, 3H). **^{13}C NMR** (125.7 MHz, DMSO-d_6): δ 174.2, 173.0, 166.4, 165.5, 156.6, 151.5, 149.6, 144.1, 142.7, 138.4, 133.6, 132.7, 131.7, 131.7, 131.7, 130.8, 130.2, 129.3, 128.9, 128.4, 127.8, 127.2, 126.9, 126.2, 126.0, 125.2, 119.1, 116.5, 116.4, 114.5, 111.7, 100.7, 85.4, 78.4, 68.0, 56.4, 34.7, 31.0, 30.8, 28.7, 28.6, 28.4, 23.6, 22.0, 13.9. **HRMS** (ESI+) *m/z* calculated for $[M+H]^+$ $C_{55}H_{49}N_4O_6$ 861.3652, found 861.3662.

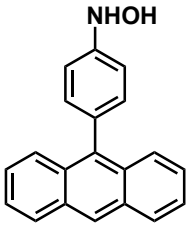
9-Nitrophenylanthracene[Ref. 339, 340], 100



4-Nitrobenzeneboronic acid (1.56 g, 9.35 mmol), 9-bromoanthracene (2.00 g, 7.78 mmol) and K_3PO_4 (4.95 g, 23.3 mmol) were dissolved in PhMe (25.0 mL), EtOH (25.0 mL) and H_2O (17.0 mL) and the mixture was degassed with argon for 30 minutes. $Pd(PPh_3)_4$ (0.449 g, 0.389 mmol) was added and the mixture degassed for further 15 minutes and heated at 110 °C overnight. The resulting mixture was diluted with CH_2Cl_2 and washed with brine. The organic fractions were dried over $MgSO_4$, filtered and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography from a mixture of hexane:CH_2Cl_2 (10 : 1) to afford the nitro compound **100** as a bright yellow crystalline solid (2.33 g, 79%). Procedure was adapted from Ref. 341.

M.p.: 227.5–229.0 °C (lit.[339] 229.0–230.0 °C). **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 8.57 (s, 1H), 8.48–8.45 (m, 2H), 8.08 (d, $J = 8.5$ Hz, 2H), 7.65–7.62 (m, 2H), 7.53–7.48 (m, 4H), 7.41–7.38 (m, $J = 1.2$ Hz, 2H). **^{13}C NMR** (125.8 MHz, $CDCl_3$) δ 147.6, 146.4, 134.1, 132.5, 131.3, 129.8, 128.8, 127.9, 126.3, 125.9, 125.5, 123.8. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{20}H_{13}NNaO_2$, 322.0838; found, 322.0834.

9-Phenylhydroxylamine anthracene, 101

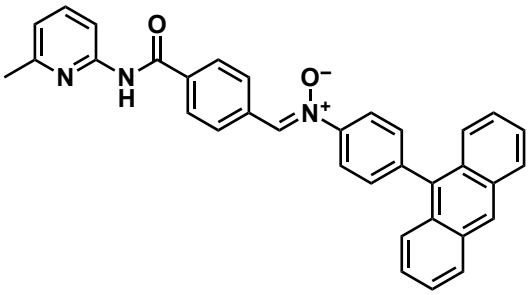


Nitro compound **100** (148 mg, 0.490 mmol) was dissolved in dry THF (15.0 mL) and Rh/C catalyst (5% wt, 60 mg) was added. Hydrazine monohydrate (0.145 mL, 0.150 mg, 2.99 mmol) was added dropwise to the suspension under argon. The reaction progress was monitored by TLC (CH_2Cl_2), staining with ninhydrin solution. Upon complete consumption of the starting material (*ca.* 20 minutes), the mixture was filtered through

celite, washing with dry THF. The solvent was removed *in vacuo* (at 25 °C, in the dark), affording the target hydroxylamine **101** as a pale yellow crystalline solid, sufficiently pure for immediate use (quantitative conversion).

^{1}H NMR (499.9 MHz, $CDCl_3$) δ 8.48 (s, 1H), 8.06–8.03 (m, 2H), 7.73–7.71 (m, 2H), 7.47–7.44 (m, 2H), 7.37–7.32 (m, 4H), 7.24–7.21 (m, 2H), 6.97 (br s, 1H), 5.30 (br s, 1H).

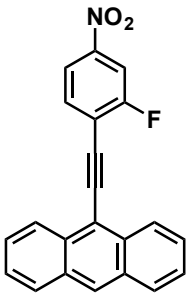
(*Z*)-*N*-(4-(Anthracen-9-yl)phenyl)-1-(4-((6-methylpyridin-2-yl)carbamoyl)phenyl)-methanimine oxide, nitrone N$^{\mathbf{A}}$



Hydroxylamine **101** (0.200 g, 0.701 mmol) and aldehyde **A** (0.150 g, 0.627 mmol) were dissolved in EtOH (14.0 mL) and $CHCl_3$ (10.0 mL) and the mixture was thoroughly sonicated. The solution was left to stand in a fridge in at 4 °C. After three days, the solution containing a yellow precipitate was filtered under suction, and the crude product was recrystallised from cold EtOAc/pet. ether 40–60 °C, furnishing the nitrone **N$^{\mathbf{A}}$** as a bright yellow solid (0.139 g, 44%).

M.p.: 254.0–255.3 °C. **^{1}H NMR** (499.9 MHz, $CDCl_3$) δ 8.62–8.54 (m, 4H), 8.23–8.22 (m, 2H), 8.11–8.07 (m, 4H), 8.05–8.02 (m, 2H), 7.69 (t, $J = 7.9$ Hz, 1H), 7.64–7.59 (m, 4H), 7.51–7.48 (m, 2H), 7.42–7.38 (m, 2H), 6.98 (d, $J = 7.5$ Hz, 1H). **^{13}C NMR** (125.7 MHz, $CDCl_3$) δ 164.8, 157.1, 150.7, 148.4, 141.6, 139.0, 135.8, 135.0, 134.1, 133.7, 132.4, 131.4, 130.1, 129.3, 128.7, 127.8, 127.5, 126.4, 126.0, 125.4, 122.0, 119.8, 111.2, 24.2. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{34}H_{26}N_3O_2$, 508.2020; found, 508.2008.

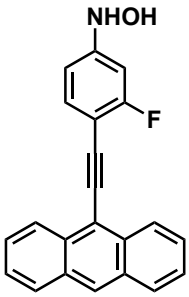
9-((2-Fluoro-4-nitrophenyl)ethynyl)anthracene, 102



Crude alkyne compound **97** (2.95 g, 14.6 mmol), 1-bromo-2-fuoro-4-nitrobenzene (1.07 g, 4.87 mmol), PPh_3 (0.128 g, 0.489 mmol) and CuI (93.0 mg, 0.488 mmol) were added to Et_3N (200 mL) and degassed with argon for 30 minutes. $Pd(PPh_3)_2Cl_2$ (0.171 g, 0.244 mmol) was added to the solution and the mixture degassed for further 10 minutes. The reaction was heated at 90 °C for three days. The crude reaction mixture was concentrated under reduced pressure and purified by flash column chromatography from a mixture of hexane:CH_2Cl_2 (6 : 1) to give the target compound **102** as a dark red crystalline solid (0.43 g, 26%).

M.p.: 210.7–213.3 °C. **^{1}H NMR** (499.9 MHz, $CDCl_3$): δ 8.61 (dd, J = 8.6, 0.8 Hz, 2H), 8.52 (s, 1H), 8.08–8.02 (m, 4H), 7.84 (dd, J = 8.6, 6.9 Hz, 1H), 7.67–7.64 (m, 2H), 7.57–7.54 (m, 2H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 162.0 (d, J = 256.1 Hz), 147.6 (d, J = 8.2 Hz), 133.4 (d, J = 1.7 Hz), 133.1, 131.2, 129.8, 129.0, 127.6, 126.4, 126.1, 119.6 (d, J = 16.4 Hz), 119.5 (d, J = 3.6 Hz), 115.5, 111.6 (d, J = 26.3 Hz), 97.4, 92.5. **^{19}F$\{^1$H$\}$ NMR** (470.4 MHz, $CDCl_3$): δ −105.3.

***N*-(4-(Anthracen-9-ylethynyl)-3-fluorophenyl)hydroxylamine, 103**

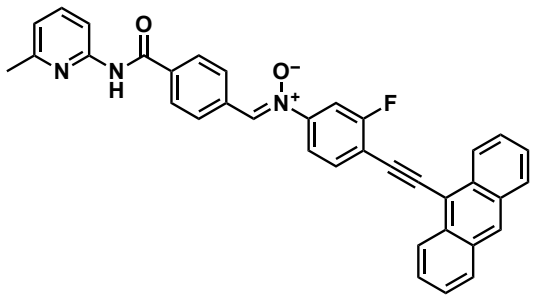


Nitro compound **102** (0.200 g, 0.586 mmol) was dissolved in dry THF (20.0 mL) and Rh/C (5% wt, 80 mg) was added. The solution was stirred for five minutes at RT before hydrazine monohydrate (0.176 µL, 0.182 g, 0.364 mmol) was added dropwise. After 25

minutes, TLC analysis (cyclohexane:EtOAc, 3 : 1) showed complete consumption of the starting material. The reaction mixture was filtered through celite, washing with dry THF and concentrated under reduced pressure (at 25 °C) to give the crude hydroxylamine **103** as a red-brown solid (quantitative conversion) that was used without purification.

^{1}H NMR (499.9 MHz, $CDCl_3$): δ 8.70–8.66 (m, 2H), 8.43 (s, 1H), 8.04–8.00 (m, 2H), 7.66–7.58 (m, 3H), 7.53–7.49 (m, 2H), 6.99–6.89 (m, 2H), 6.79 (dd, J = 8.2, 2.1 Hz, 1H), 5.22 (br s, 1H). **^{19}F$\{^1$H$\}$ NMR** (470.4 MHz, $CDCl_3$): δ −107.7.

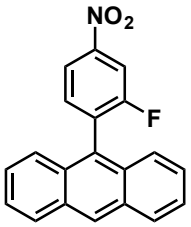
(*Z*)-*N*-(4-(anthracen-9-ylethynyl)-3-fluorophenyl)-1-(4-((6-methylpyridin-2-yl)-carbamoyl)phenyl)methanimine oxide, nitrone N^{EF}



Hydroxylamine **103** (0.190 g, 0.581 mmol) and aldehyde **A** (0.126 g, 0.523 mmol) were dissolved in a mixture of EtOH (14.0 mL) and $CHCl_3$ (14.0 mL) and sonicated thoroughly. The reaction mixture was left unstirred for three days at 6 °C. After this time, formation of red precipitate was visible and the solid was isolated by filtration under suction, using pet. ether 40–60 °C to wash the product, giving the desired nitrone **N^{EF}** as a red solid (107 g, 37%).

M.p.: 214.4–216.0 °C. **^{1}H NMR** (499.9 MHz, $CDCl_3$): δ 8.67 (dd, J = 8.7, 0.9 Hz, 2H), 8.59 (s, 1H), 8.57–8.54 (m, 2H), 8.51 (s, 1H), 8.20 (d, J = 8.2 Hz, 1H), 8.10 (s, 1H), 8.08–8.04 (m, 4H), 7.88 (dd, J = 8.3, 7.2 Hz, 1H), 7.79 (dd, J = 9.4, 2.2 Hz, 1H), 7.71–7.64 (m, 4H), 7.57–7.54 (m, J = 1.3 Hz, 2H), 6.97 (d, J = 7.5 Hz, 1H), 2.50 (s, 3H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 164.7, 162.7 (d, J = 255.1 Hz), 157.2, 150.7, 149.2 (d, J = 8.4 Hz), 139.0, 136.2, 133.7, 133.6, 133.6, 133.0, 131.3, 129.5, 129.1, 129.0, 127.8, 127.3, 126.6, 126.0, 119.9, 117.3 (d, J = 3.6 Hz), 116.2, 115.1, 111.2, 110.2 (d, J = 26.3 Hz), 94.5, 92.8, 24.2. **^{19}F$\{^1$H$\}$ NMR** (470.4 MHz, $CDCl_3$): δ −105.7.

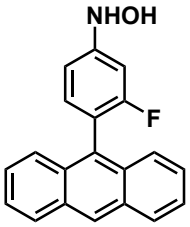
9-(2-Fluoro-4-nitrophenyl)anthracene[Ref. 334], 104



9-Anthraceneboronic acid (1.07 g, 4.80 mmol) and 1-bromo-2-fluoro-4-nitrobenzene (0.880 g, 4.00 mmol) were added to a suspension of K_3PO_4 (2.55 g, 12.0 mmol) in a mixture of PhMe (13.5 mL), EtOH (13.5 mL) and H_2O (9.00 mL). The mixture was vigorously degassed with argon for 20 minutes. $Pd(PPh_3)_4$ (0.231 g, 0.200 mmol) was added under the flow of nitrogen and the suspension degassed for 10 more minutes. The reaction was heated at 100 °C for 16 hours under inert atmosphere. The organic product was extracted with CH_2Cl_2 and washed with brine. The crude product was purified by flash column chromatography, employing hexane:CH_2Cl_2 (3 : 1) as the eluent. The product **104** was isolated as an orange solid (1.20 g, 77%). Procedure was adapted from Ref. 341.

M.p.: 179.6–181.4 °C. (lit.[334] 182–183 °C). **^{1}H NMR** (499.9 MHz, $CDCl_3$): δ 8.61 (s, 1H), 8.28 (dd, $J = 8.3$, 1.7 Hz, 1H), 8.22 (dd, $J = 8.5$, 2.1 Hz, 1H), 8.10 (d, $J = 8.1$ Hz, 2H), 7.62 (dd, $J = 8.1$, 7.1 Hz, 1H), 7.53–7.50 (m, 4H), 7.45–7.42 (m, 2H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 160.4 (d, $J = 253.0$ Hz), 148.3 (d, $J = 8.0$ Hz), 134.4 (d, $J = 4.0$ Hz), 133.8 (d, $J = 18.0$ Hz), 131.3, 130.0, 128.9, 128.8, 127.5, 126.7, 125.6, 125.3, 119.5 (d, $J = 4.0$ Hz), 112.1 (d, $J = 28.0$ Hz). **^{19}F{^{1}H} NMR** (470.4 MHz, $CDCl_3$): δ −108.0. **HRMS** (CI+): *m/z* calculated for $[M+H]^+$ $C_{20}H_{13}FNO_2$, 318.0925; found, 318.0929.

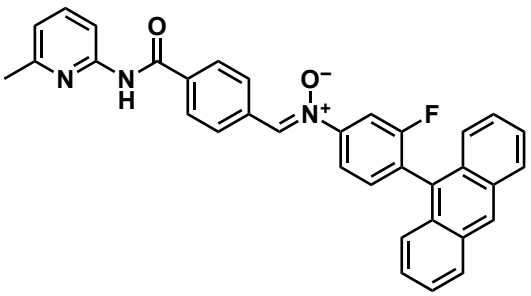
***N*-(4-(Anthracen-9-yl)-3-fluorophenyl)hydroxylamine, 105**



Nitro compound **104** (0.300 g, 0.945 mmol) was dissolved in dry THF (30.0 mL) and Rh/C (5% wt, 120 mg) was added. The solution was stirred for five minutes at RT before hydrazine monohydrate (0.135 mL, 0.139 mg, 2.78 mmol) was added dropwise. After 35 minutes, TLC analysis (hexane:EtOAc, 3 : 1) showed complete consumption of the starting material. The reaction mixture was filtered through celite and concentrated under reduced pressure (at 25 °C) to give the crude hydroxylamine **105** as a yellow solid (quantitative conversion) which was used without purification.

^{1}H NMR (499.9 MHz, $CDCl_3$): δ 8.52 (s, 1H), 8.07–8.04 (m, 2H), 7.69–7.67 (m, 2H), 7.48–7.45 (m, 2H), 7.41–7.37 (m, 2H), 7.29–7.27 (m, 1H), 7.05–7.02 (m, 1H), 7.01–6.97 (m, 1H), 6.96–6.93 (m, 1H), 5.33 (br s, 1H). **^{19}F{^{1}H} NMR** (470.4 MHz, $CDCl_3$): δ −112.0. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{20}H_{13}FNO$, 302.0987; found, 302.0983.

(*Z*)-*N*-(4-(Anthracen-9-yl)-3-fluorophenyl)-1-(4-((6-methylpyridin-2-yl)carbamoyl)phenyl)methanimine oxide, nitrone N^{AF}

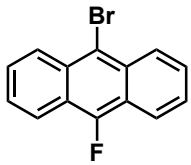


The hydroxylamine **105** (0.200 g, 0.659 mmol) and aldehyde **A** (0.142 g, 0.593 mmol) were dissolved in EtOH (14.0 mL) and $CHCl_3$ (10.0 mL) and the solution was thoroughly sonicated and left in the fridge (at 6 °C) for three days. The yellow precipitate was filtered under suction, affording the crude nitrone that was purified *via* flash column chromatography (CH_2Cl_2 to CH_2Cl_2:EtOAC, 1 : 1), the fractions containing the product were concentrated under a flow of compressed air. Pet. ether 40–60 °C was added to

precipitate the nitrone $\mathbf{N^{AF}}$, which was isolated by vacumm filtration as a pale yellow solid (0.128 g, 41%).

M.p.: 267.3–268.4 °C. **^{1}H NMR** (499.9 MHz, $CDCl_3$): δ 8.61–8.56 (m, 4H), 8.22–8.21 (m, 2H), 8.12–8.08 (m, 4H), 7.88 (dd, $J = 9.1$, 2.1 Hz, 1H), 7.83 (dd, $J = 8.1$, 2.1 Hz, 1H), 7.70–7.67 (m, 1H), 7.62–7.55 (m, 3H), 7.53–7.49 (m, 2H), 7.47–7.42 (m, 2H), 6.98 (d, $J = 7.3$ Hz, 1H), 2.51 (s, 3H). **^{19}F{^{1}H} NMR** (470.4 MHz, $CDCl_3$): δ −109.0. **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 164.7, 160.7 (d, $J = 250.5$ Hz), 157.2, 150.7, 149.9 (d, $J = 8.3$ Hz), 139.0, 136.1, 134.2 (d, $J = 4.1$ Hz), 133.8, 133.7, 131.4, 130.3, 129.4, 128.9, 128.8, 128.7, 128.3, 127.8, 126.4, 125.7, 125.5, 119.8, 117.4 (d, $J = 3.5$ Hz), 111.2, 110.6 (d, $J = 27.1$ Hz), 24.2. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{34}H_{25}FN_3O_2$, 526.1925; found, 526.1915.

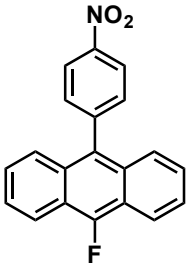
9-Bromo-10-fluoroanthracene, 106



n-Butyllithium (15.0 mL, 2.22 M, 33.3 mmol) was slowly injected to an ice-cold solution of 9,10-dibromoanthracene (predried over P_2O_5 under vacumm for one week) in Et_2O (150 mL) under inert atmosphere and the mixture was stirred at 0 °C for a period of 30 minutes. *N*-fluorobenzenesulfonimide (12.1 g, 38.3 mmol, predried over P_2O_5 under vacumm for one week) was added portionwise to the solution, and the reaction mixture was allowed to warm to RT. The reaction was heated at 30 °C overnight. The reaction was quenched by addition of H_2O (5.00 mL). Solvent was removed under reduced pressure and the residue was redissolved in CH_2Cl_2 and washed with H_2O. The *aq.* fractions were washed with CH_2Cl_2 and the combined organic fractions were washed with brine. The solution was dried with $MgSO_4$, filtered and the solvent removed under reduced pressure. The crude product was purified by column chromatography (pet. ether 40–60 °C) followed by recrystallisation from hot EtOH (3 repeats), affording the product **106** as clear yellow needles, containing a mixture of the desired product, 9-bromo-10-fluoroanthracene **106** (86%) and 9,10-dibromoanthracene (14%), and was used without further purification (1.55 g, 17%).

^{1}H NMR (400.3 MHz, $CDCl_3$): δ 8.52–8.49 (m, 2H), 8.31–8.28 (m, 2H), 7.65–7.61 (m, 2H), 7.58–7.54 (m, 2H). **^{19}F$\{^{1}$H$\}$ NMR** (376.6 MHz, $CDCl_3$): δ −130.0. **HRMS** (ESI+): *m/z* calculated for $[M]^+$ $C_{14}H_8BrF$, 273.9793; found, 273.9787.

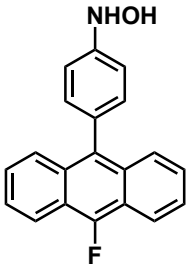
9-fluoro-10-(4-nitrophenyl)anthracene, 107



Compound **106** (0.800 g, *ca.* 2.91 mmol), nitrophenylboronic acid (0.631 g, 3.78 mmol) and K_3PO_4 (1.85 g, 8.72 mmol) were dissolved in a mixture of EtOH (12.0 mL), PhMe (12.0 mL) and H_2O (8.0 mL) and the solution was degassed with argon for 30 minutes. $Pd(PPh_3)_4$ was added and the mixture was degassed for further 15 minutes. The reaction mixture was heated at 100 °C overnight. The mixture was diluted with CH_2Cl_2 and washed with brine, dried with $MgSO_4$, filtered and the solvent removed under reduced pressure. The crude product was purified by flash chromatography (pet. ether 40–60 °C:CH_2Cl_2, 5 : 1) to afford the desired product **107** as a bright yellow solid (0.613 g, 66%). Procedure was adapted from Ref. 341.

M.p.: 224.7–227.4 °C. **^{1}H NMR** (400.1 MHz, $CDCl_3$): δ 8.48–8.44 (m, 2H), 8.38–8.35 (m, 2H), 7.63–7.60 (m, 2H), 7.57–7.51 (m, 4H), 7.45–7.41 (m, 2H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 154.6 (d, J = 259.0 Hz), 147.7, 145.9, 132.8, 132.7, 130.2 (d, J = 5.0 Hz), 129.5 (d, J = 8.0 Hz), 126.8 (d, J = 1.7 Hz), 125.7 (d, J = 1.7 Hz), 125.7, 123.9, 120.9 (d, J = 5.8 Hz), 118.8 (d, J = 15.0 Hz). **^{19}F$\{^{1}$H$\}$ NMR** (376.5 MHz, $CDCl_3$): δ −128.9. **HRMS** (ESI+): *m/z* calculated for $[M]^+$ $C_{20}H_{12}FNO_2$, 317.0852; found, 317.0846.

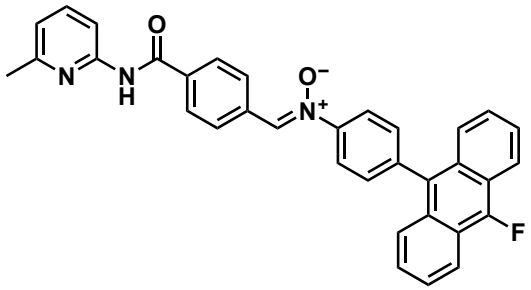
***N*-(4-(10-fluoroanthracen-9-yl)phenyl)hydroxylamine, 108**



Nitro compound **107** (0.200 g, 0.630 mmol) was dissolved in dry THF (20.0 mL) and Rh/C (5% wt, 80 mg) was added. The suspension was stirred for five minutes under argon balloon before hydrazine monohydrate (0.126 mL, 0.130 mg, 2.52 mmol) was added dropwise. The reaction was monitored by TLC (pet. ether 40–60 °C:CH_2Cl_2, 5 : 3). Upon completion (*ca.* 20 minutes), the reaction mixture was filtered through celite, washing with dry THF. The solvent was removed under reduced pressure at 25 °C, affording the crude hydroxylamine **108** as a pale yellow solid, sufficiently pure for immediate use in the next step (quantitative conversion).

^{1}H NMR δ (500.1 MHz, $CDCl_3$): δ 8.34–8.31 (m, 2H), 7.73–7.70 (m, 2H), 7.53–7.48 (m, 2H), 7.40–7.36 (m, 2H), 7.34–7.31 (m, 2H), 7.23–7.20 (m, 2H), 6.97 (br s, 1H), 5.36 (br s, 1H). **^{19}F$\{^1$H$\}$ NMR** (470.4 MHz, $CDCl_3$): δ −131.5. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{20}H_{13}FNO$, 302.0987; found, 302.0990.

(*Z*)-*N*-(4-(10-Fluoroanthracen-9-yl)phenyl)-1-(4-((6-methylpyridin-2-yl)carbamoyl)phenyl)methanimine oxide, nitrone N^{A10F}

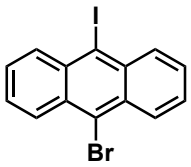


Hydorxylamine **108** (0.190 g, 0.626 mmol) and aldehyde **A** (0.136 g, 0.564 mmol) were dissolved in EtOH (10.0 mL) and $CHCl_3$ (5.00 mL) and sonicated thoroughly. The reaction mixture was kept at 5 °C for four days, the resulting precipitate was filtered under suction and washed with pet. ether 40–60 °C. During the filtration, formation of further precipitate was observed in the filtrate, which was again isolated by vacuum filtration.

The second precipitate afforded the product $\mathbf{N^{A10F}}$ as a pale yellow solid in sufficient purity for further reactions (0.102 g, 34%).

M.p.: 232.4–235.1 °C. **^{1}H NMR** (400.1 MHz, $CDCl_3$): δ 8.61–8.58 (m, 3H), 8.38–8.34 (m, 2H), 8.23–8.20 (m, 2H), 8.11–8.07 (m, 2H), 8.05–8.01 (m, 2H), 7.68 (t, $J = 7.8$ Hz, 1H), 7.64–7.60 (m, 2H), 7.59–7.52 (m, 4H), 7.45–7.40 (m, 2H), 6.98–6.96 (m, 1H), 2.51 (s, 3H). **^{13}C NMR** (100.7 MHz, $CDCl_3$): δ 164.8, 157.2, 154.4 (d, $J = 258.2$ Hz), 150.8, 148.5, 141.1, 139.0, 135.9, 134.0, 133.7, 132.6, 130.6, 130.5, 129.3, 127.8, 126.5 (d, $J = 1.1$ Hz), 126.1 (d, $J = 3.2$ Hz), 125.6 (d, $J = 3.1$ Hz), 122.2 (d, $J = 47.7$ Hz), 120.8 (d, $J = 5.70$ Hz), 119.8, 118.9 (d, $J = 14.5$ Hz), 111.2, 24.2. **^{19}F$\{^1$H$\}$ NMR** (376.6 MHz, $CDCl_3$): δ −129.8. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{34}H_{25}FN_3O_2$, 526.1925; found, 526.1922.

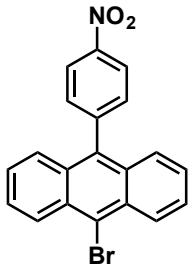
9-Bromo-10-iodoanthracene[Ref. 342], 109



9,10-bromoanthracene (10.0 g, 29.8 mmol) was added to dry THF (250 mL) in a flame-dried flask and the solution was degassed vigorously with argon for 20 minutes. The reaction mixture was cooled to −70 °C using a dry ice/acetone bath and *n*-butyllithium (17.8 mL, 1.85 M, 32.9 mmol) was added dropwise. The resulting red solution was stirred at −70 °C for three hours, at which point I_2 (9.75 g, 38.4 mmol) dissolved in dry THF (3.0 mL) was added dropwise. After one hour, the reaction was allowed to slowly warm to RT overnight. The solution was concentrated to 10% of it volume under reduced pressure and *aq.* solution of $Na_2S_2O_3$ (20% wt, 150 mL) was added. The resulting precipitate was filtered under suction, and washed with further solution of $Na_2S_2O_3$. The crude product was recrystallised from hot PhMe to give the desired mono-iodinated product **109** as yellow crystalline solid (5.62 g, 49%) which contained some residual starting material and 9,10-diodoanthracene (*ca.* 10%), and was used without further purification.

^{1}H NMR (500.1 MHz, $CDCl_3$): δ 8.59–8.55 (m, 4H), 7.64–7.58 (m, 4H). **HRMS** (CI−): *m/z* calculated for $[M]^-$ $C_{14}H_8BrI$, 381.8854; found, 381.8862.

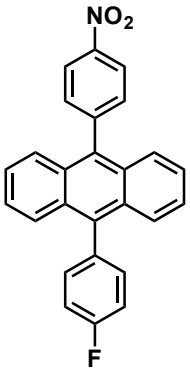
9-Bromo-10-(4-nitrophenyl)anthracene, 110



9-Bromo-10-iodoanthracene **109** (0.500 g, 1.31 mmol) and 4-nitrobenzeneboronic acid (0.262 g, 1.57 mmol) were added with Na_2CO_3 (0.207 g, 1.95 mmol), dppf (72.3 mg, 0.131 mmol) and $Pd(OAc)_2$ (29.3 mg, 0.130 mmol) to a mixture of THF (12.5 mL) and H_2O (7.50 mL). The mixture was heated in a CEM microwave at 90 °C for 150 minutes. Once finished, the crude reaction mixture was diluted with CH_2Cl_2/H_2O and the product was extracted into CH_2Cl_2 and the combined organic fractions were washed with brine and dried over $MgSO_4$, filtered and the solvent removed under reduced pressure. The crude product was purified using flash column chromatography (hexane:CH_2Cl_2, 3 : 1) to afford the intermediate nitro compound **110** as a bright yellow crystalline solid (0.55 g, 56%).

M.p.: 276.1–278.8 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 8.65 (d, J = 8.9 Hz, 2H), 8.48–8.46 (m, 2H), 7.64–7.60 (m, 4H), 7.50 (d, J = 8.9 Hz, 2H), 7.44–7.41 (m, 2H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 147.8, 146.0, 134.8, 132.4, 130.5, 130.3, 128.3, 127.3, 126.5, 126.5, 124.3, 123.9. **HRMS** (CI+): *m/z* calculated for $[M+H]^+$ $C_{20}H_{13}BrNO_2$, 378.0124; found, 378.0132.

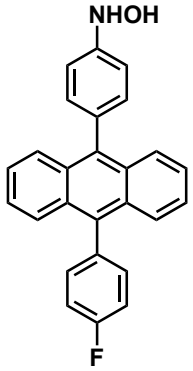
9-(4-Fluorophenyl)-10-(4-nitrophenyl)anthracene, 111



9-bromo-10-(4-nitrophenyl)anthracene **110** (0.400 g, 1.06 mmol), 4-fluorobenzeneboronic acid (0.192 g, 1.37 mmol), Na_2CO_3 (0.224 g, 2.11 mmol), dppf (87.8 mg, 0.158 mmol) and $Pd(OAc)_2$ (35.5 mg, 0.158 mmol) were suspended in dioxane (11.0 mL) and H_2O (7.00 mL). The mixture was heated in a CEM microwave at 110 °C for four hours. Once finished, the crude reaction mixture was diluted with CH_2Cl_2/H_2O and the product was extracted into CH_2Cl_2 and the combined organic fractions were washed with brine and dried over $MgSO_4$, filtered and the solvent removed under reduced pressure. The crude product was purified using flash column chromatography (hexane:CH_2Cl_2, 1 : 1) to afford the final nitro compound **111** as a bright yellow crystalline solid (0.343 g, 82%).

M.p.: 358.3–360.6 °C. **^{1}H NMR** (400.3 MHz, $CDCl_3$): δ 8.51–8.48 (m, 2H), 7.72–7.67 (m, 4H), 7.57–7.53 (m, 2H), 7.46–7.43 (m, 2H), 7.41–7.36 (m, 4H), 7.34–7.30 (m, 2H). **^{19}F{^{1}H} NMR** (376.6 MHz, $CDCl_3$): δ −114.5. **^{13}C NMR** (100.7 MHz, $CDCl_3$): δ 162.6 (d, J = 247.0 Hz), 147.7, 146.7, 137.3, 134.5, 134.5, 132.9 (d, J = 7.7 Hz), 132.6, 130.1, 129.5, 127.2, 126.2, 126.0, 125.8, 123.9, 115.7 (d, J = 21.4 Hz). **HRMS** (CI+): *m/z* calculated for $[M+H]^+$ $C_{26}H_{17}FNO_2$, 394.1238; found, 394.1244.

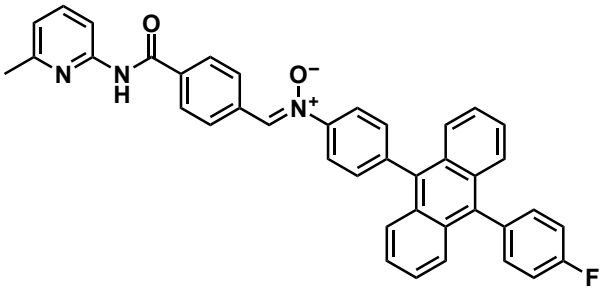
***N*-(4-(10-(4-Fluorophenyl)anthracen-9-yl)phenyl)hydroxylamine, 112**



Nitro compound **111** (0.100 g, 0.254 mmol) was dissolved in dry THF (10.0 mL) and Rh/C (5% wt, 40 mg) was added. Once dissolved, hydrazine monohydrate (40.0 µL, 41.3 µg, 0.825 mmol) was added dropwise to the suspension under argon atmosphere. The reaction progress was monitored by TLC (hexane:CH_2Cl_2, 1 : 1) and staining with ninhydrin dye. Upon completion (*ca.* 30 minutes), the reaction mixture was filtered through celite (with dry THF) and the solvent was removed under reduced pressure (at 25 °C) to yield hydroxylamine product **112** as a pale yellow solid, sufficiently pure for immediate use in further reactions (quantitative conversion).

^{1}H NMR (499.9 MHz, $CDCl_3$): δ 7.79–7.77 (m, 2H), 7.68–7.66 (m, 2H), 7.49–7.44 (m, 2H), 7.42–7.40 (m, 2H), 7.39–7.25 (m, 8H), 7.01 (br s, 1H), 5.34 (br s, 1H). **^{19}F$\{^1$H$\}$ NMR** (470.3 MHz, $CDCl_3$): δ −115.1.

(*Z*)-*N*-(4-(10-(4-Fluorophenyl)anthracen-9-yl)phenyl)-1-(4-((6-methylpyridin-2-yl)-carbamoyl)phenyl)methanimine oxide, nitrone N^{AFPh}

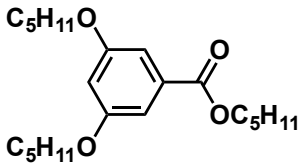


Hydroxylamine **112** (88.5 mg, 0.233 mmol) and aldehyde **A** (50.4 mg, 0.210 mmol) were dissolved in EtOH (2.00 mL) and CH_2Cl_2 (7.00 mL). The solution was thoroughly sonicated and stirred at RT overnight, then left to stand at −20 °C overnight. The resulting precipitate was filtered, washing with pet. ether 40–60 °C. The crude nitrone was

purified by flash chromatography (CH_2Cl_2 to EtOAc) to afford the final nitrone $\mathbf{N^{AFPh}}$ as an orange (yellow fluorescent) solid, in quantity sufficient only for characterisation by **^{1}H NMR** and **^{19}F{^{1}H} NMR** spectroscopy.

^{1}H NMR (499.9 MHz, $CDCl_3$): δ 8.64 (s, 1H), 8.61–8.59 (m, 2H), 8.24–8.21 (m, 2H), 8.10–8.08 (m, 2H), 8.07–8.04 (m, 2H), 7.71–7.62 (m, 7H), 7.47–7.43 (m, 2H), 7.40–7.37 (m, 4H), 7.34–7.30 (m, 2H), 6.97 (d, $J = 7.5$ Hz, 1H), 2.51 (s, 3H). **^{19}F{^{1}H} NMR** (470.4 MHz, $CDCl_3$): δ −114.7.

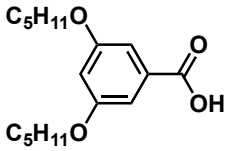
Pentyl 3,5-bis(pentyloxy)benzoate, 158



3,5-dihydroxybenzoic acid (5.00 g, 32.4 mmol), KI (5.30 g, 31.9 mmol) and K_2CO_3 (27.0 g, 195 mmol) were suspended in dry MeCN (80.0 mL) and dry DMF (80.0 mL). Bromopentane (29.0 g, 194 mmol) was injected and the reaction was heated to 92 °C for 16 hours, or until TLC analysis (hexane:EtOAc, 2 : 1) showed complete consumption of the 3,5-dihydroxybenzoic acid. The reaction mixture was cooled to RT and MeCN was removed under reduced pressure. The mixture was neutralised with *aq.* HCl (1 M) and diluted with H_2O. The product was extracted with EtOAc and the combined organic fractions were washed with a *sat.* LiCl. The solution was dried over $MgSO_4$, filtered and solvent was removed under reduced pressure to yield the crude product, which was purified by flash chromatography (hexane:EtOAc, 2 : 1) to afford ester **158** as a pale yellow oil (10.6 g, 90%).

^{1}H NMR (400.1 MHz, $CDCl_3$): δ 7.16 (d, $J = 2.4$ Hz, 2H), 6.63 (t, $J = 2.4$ Hz, 1H), 4.29 (t, $J = 6.7$ Hz, 2H), 3.97 (t, $J = 6.6$ Hz, 4H), 1.82–1.72 (m, 6H), 1.48–1.34 (m, 12H), 0.96–0.90 (m, 9H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 166.7, 160.2, 132.3, 107.7, 106.3, 68.4, 65.4, 29.0, 28.5, 28.3, 28.3, 22.6, 22.5, 14.1, 14.1. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{22}H_{36}NaO_4$ 387.2506, found 387.2496.

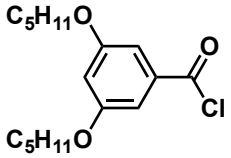
3,5-Bis(pentyloxy)benzoic acid, 159



Ester **158** (11.8 g, 32.4 mmol) was dissolved in dry THF (60.0 mL) and *aq.* KOH (10 M, 25.0 mL). The solution was heated to 82 °C overnight, or until TLC analysis (hexane:EtOAc, 2 : 1) showed complete consumption of the starting material. The reaction mixture was allowed to cool to RT and the THF was removed under reduced pressure. The remaining solution was carefully acidified to pH 1 with *aq.* HCl (1 M) and the product was extracted with CH_2Cl_2. The organic fractions were combined, washed with brine and dried over $MgSO_4$. The mixture was filtered and solvent was removed *in vacuo*. The crude product was purified by flash chromatography (hexane:EtOAc, 2 : 1) to afford the desired product **159** as an off-white solid (8.00 g, 80%).

M.p.: 62.7–64.4 °C. **^{1}H NMR** (400.3 MHz, $CDCl_3$): δ 7.23 (d, $J = 2.3$ Hz, 2H), 6.69 (t, $J = 2.3$ Hz, 1H), 3.98 (t, $J = 6.6$ Hz, 4H), 1.82–1.77 (m, 4H), 1.48–1.35 (m, 8H), 0.94 (t, $J = 7.2$ Hz, 6H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 172.0, 160.3, 131.0, 108.3, 107.6, 68.5, 29.0, 28.3, 22.6, 14.2. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{17}H_{26}NaO_4$ 317.1723, found 317.1718.

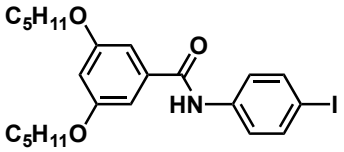
3,5-Bis(pentyloxy)benzoyl chloride, 160



3,5-Bis(pentyloxy)benzoic acid **159** (4.00 g, 13.0 mmol) was dissolved in dry PhMe (80.0 mL) under argon and $SOCl_2$ (8.00 mL, 110 mmol) was injected dropwise. The solution was heated to 95 °C for 16 hours, or until TLC analysis (hexane:EtOAc, 2 : 1) showed complete consumption of the 3,5-bis(pentyloxy)benzoic acid **159**. The mixture was allowed to cool to RT and solvent was removed under reduced pressure to afford the crude activated acyl chloride **160** as a brown oil that was used immediately without further purification (quantitative conversion).

^{1}H NMR (400.1 MHz, $CDCl_3$): δ 7.22 (d, J = 2.3 Hz, 2H), 6.74 (t, J = 2.3 Hz, 1H), 3.98 (t, J = 6.5 Hz, 4H), 1.84–1.77 (m, 4H), 1.49–1.36 (m, 8H), 0.94 (t, J = 7.1 Hz, 6H).

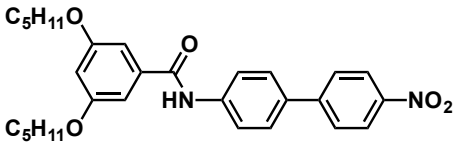
***N*-(4-Iodophenyl)-3,5-bis(pentyloxy)benzamide, 161**



The crude acyl chloride **160** was dissolved in dry CH_2Cl_2 (15.0 mL) and added dropwise to an ice-cold solution of 4-iodoaniline (4.20 g, 19.5 mmol) and Et_3N (5.00 mL, 35.8 mmol) in dry CH_2Cl_2 (40.0 mL) under argon atmosphere. Once the addition was complete, the reaction mixture was allowed to warm to RT and was stirred for 16 hours. The mixture was washed with *aq.* HCl (1 M) and brine. The organic layer was dried over $MgSO_4$, filtered and solvent was removed under reduced pressure. The crude product was purified by flash chromatography (pet. ether 40–60 °C:EtOAC, 5 : 3). All fractions containing product were combined and the solvent volume was reduced to less than 5% under reduced pressure. The desired product **161** was obtained as beige solid by precipitation by addition of petroleum ether (40–60 °C) (4.80 g, 74% over two steps).

M.p.: 89.2–89.8 °C. **^{1}H NMR** (400.1 MHz, $CDCl_3$): δ 7.73 (br s, 1H), 7.68–7.65 (m, 2H), 7.44–7.40 (m, 2H), 6.93 (d, J = 2.2 Hz, 2H), 6.61 (t, J = 2.2 Hz, 1H), 3.98 (t, J = 6.6 Hz, 4H), 1.83–1.76 (m, 4H), 1.48–1.34 (m, 8H), 0.94 (t, J = 7.1 Hz, 6H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 165.7, 160.7, 138.1, 137.8, 136.8, 122.0, 105.4, 104.8, 87.8, 68.5, 29.0, 28.3, 22.6, 14.2. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{23}H_{29}INO_3$ 494.1198 found 494.1202.

***N*-(4'-Nitro-[1,1'-biphenyl]-4-yl)-3,5-bis(pentyloxy)benzamide, N**

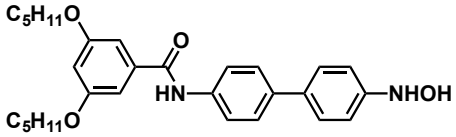


Iodo compound **161** (3.00 g, 6.06 mmol), 4-nitrobenzeneboronic acid (1.21 g, 7.27 mmol) and K_3PO_4 (3.86 g, 18.2 mmol) were dissolved in EtOH (13.0 mL), PhMe (24.0 mL) and H_2O (13.0 mL), and the mixture was degassed with argon for 30 minutes. $Pd(PPh_3)_4$ (350 mg, 0.302 mmol) was added and the mixture was degassed

for further 15 minutes, and subsequently heated to 100 °C under nitrogen until TLC analysis (CH_2Cl_2) showed complete consumption of iodo compound **161** (*ca.* 18 hours). The reaction mixture was diluted with CH_2Cl_2 and washed with brine. The organic fractions were dried over $MgSO_4$, filtered under gravity and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (CH_2Cl_2:hexane, 1 : 1) to yield the desired product **N** as a pale yellow solid (1.77 g, 60%).

M.p.: 180.6–182.4 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 8.30–8.28 (m, 2H), 7.96 (br s, 1H), 7.80–7.77 (m, 2H), 7.74–7.72 (m, 2H), 7.66–7.63 (m, 2H), 6.97 (d, $J = 2.2$ Hz, 2H), 6.62 (t, $J = 2.2$ Hz, 1H), 3.99 (t, $J = 6.6$ Hz, 4H), 1.82–1.77 (m, 4H), 1.48–1.35 (m, 8H), 0.94 (t, $J = 7.2$ Hz, 6H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 165.8, 160.7, 147.0, 146.9, 138.9, 136.8, 134.7, 128.2, 127.5, 124.3, 120.6, 105.5, 104.8, 68.5, 29.0, 28.3, 22.6, 14.2. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{29}H_{33}N_2O_5$ 489.2395, found 489.2397.

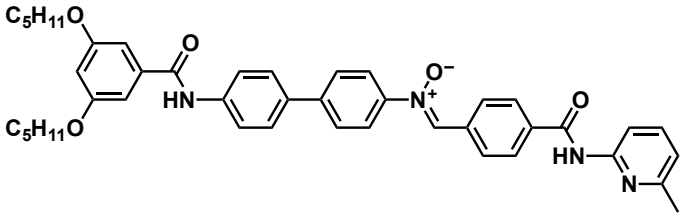
***N*-(4'-(Hydroxyamino)-[1,1'-biphenyl]-4-yl)-3,5-bis(pentyloxy)benzamide, 162**



Nitro compound **N** (200 mg, 0.408 mmol) was dissolved in dry THF (20.0 mL) and Rh/C (5% wt, 80.0 mg) was added. Hydrazine monohydrate (0.127 mL, 0.131 mg, 2.62 mmol) was added dropwise and the reaction stirred at RT until TLC analysis (pet. ether 40–60 °C:EtOAc, 5 : 4, staining with ninhydrin solution) showed complete conversion to the product (approximately 40 minutes). The reaction mixture was filtered through celite, washing with dry THF. Solvent was removed under reduced pressure (at 25 °C, in the dark) to yield the crude hydroxylamine **162** as a yellow solid, which was used without further purification (quantitative conversion).

^{1}H NMR (400.3 MHz, $CDCl_3$): δ 7.82 (br s, 1H), 7.69–7.66 (m, 2H), 7.58–7.54 (m, 2H), 7.54–7.50 (m, 2H), 7.08–7.05 (m, 2H), 6.97 (d, $J = 2.2$ Hz, 2H), 6.84 (br s, 1H), 6.62 (t, $J = 2.2$ Hz, 1H), 5.40 (br s, 1H), 4.00 (t, $J = 6.6$ Hz, 4H), 1.84–1.77 (m, 4H), 1.49–1.35 (m, 8H), 0.94 (t, $J = 7.1$ Hz, 6H). **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{29}H_{36}N_2NaO_4$ 499.2567, found 499.2562.

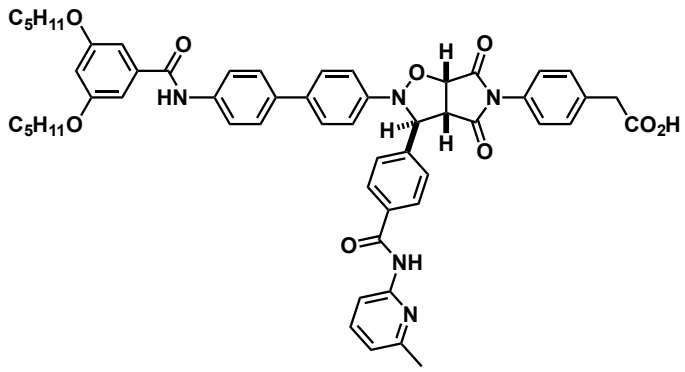
(*Z*)-*N*-(4'-(3,5-bis(pentyloxy)benzamido)-[1,1'-biphenyl]-4-yl)-1-(4-((6-methylpyridin-2-yl)carbamoyl)phenyl)methanimine oxide, linear component L



Aldehyde **A** (79.0 mg, 0.329 mmol) and hydroxylamine **162** (174 mg, 0.366 mmol) were dissolved in EtOH (9.50 mL) and $CDCl_3$ (4.50 mL). The reaction mixture was left unstirred in a fridge at 4 °C for four days, during which time a precipitate was formed. The precipitate was filtered, washing with hexane, affording the desired Linear component **L** as a bright yellow solid (141 mg, 61%).

M.p.: 207.0–208.7 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 8.57 (br s, 1H), 8.54–8.52 (m, 2H), 8.20 (d, $J = 8.2$ Hz, 1H), 8.07 (s, 1H), 8.05–8.03 (m, 2H), 7.94 (br s, 1H), 7.88–7.86 (m, 2H), 7.78–7.75 (m, 2H), 7.72–7.70 (m, 2H), 7.68–7.62 (m, 3H), 6.98 (d, $J = 2.2$ Hz, 2H), 6.96 (d, $J = 7.5$ Hz, 1H), 6.62 (t, $J = 2.2$ Hz, 1H), 4.00 (t, $J = 6.6$ Hz, 4H), 2.49 (s, 3H), 1.83–1.77 (m, 4H), 1.48–1.35 (m, 8H), 0.94 (t, $J = 7.2$ Hz, 6H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 165.8, 164.8, 160.7, 157.1, 150.7, 147.9, 142.8, 139.0, 138.2, 136.9, 135.7, 135.5, 134.1, 133.3, 129.2, 128.0, 127.7, 127.6, 122.3, 120.6, 119.8, 111.1, 105.5, 104.8, 68.5, 29.0, 28.3, 24.2, 22.6, 14.2. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{43}H_{45}N_4O_5$ 697.3395, found 697.3404. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{43}H_{47}N_4O_5$ 699.3541, found 699.3526.

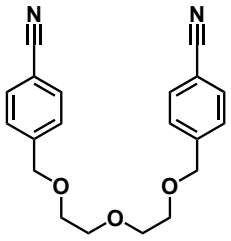
Thread, T



Linear component **L** (41.3 mg, 59.1 µmol) and maleimide stopper **S** (13.7 mg, 59.1 µmol) were weighed out in a vial and $CDCl_3$ (6.00 mL) was added and the reaction mixture was kept at 4 °C for two days at which point **^{1}H NMR** spectroscopy confirmed complete consumption of the starting material. Solvent was removed under a flow of compressed air. The crude material was purified by reverse phase chromatography (10 g C18 column), eluting with 80% MeCN and 20% H_2O, to give the desired thread **T** as an off-white powdery solid (25.0 mg, 45%).

M.p.: $>$ 205 °C (decomp.). **^{1}H NMR** (499.9 MHz, DMSO-d_6): δ 12.39 (br s, 1H), 10.78 (s, 1H), 10.22 (s, 1H), 8.11–8.08 (m, 2H), 8.03 (d, $J = 8.2$ Hz, 1H), 7.86–7.84 (m, 2H), 7.75–7.70 (m, 3H), 7.63–7.58 (m, 4H), 7.34–7.32 (m, 2H), 7.18–7.16 (m, 2H), 7.10 (d, $J = 2.2$ Hz, 2H), 7.04 (d, $J = 7.5$ Hz, 1H), 6.69 (t, $J = 2.2$ Hz, 1H), 6.56–6.54 (m, 2H), 6.07 (s, 1H), 5.45 (d, $J = 7.5$ Hz, 1H), 4.20 (d, $J = 7.5$ Hz, 1H), 4.03 (t, $J = 6.4$ Hz, 4H), 3.55 (s, 2H), 2.46 (s, 3H), 1.76–1.70 (m, 4H), 1.45–1.32 (m, 8H), 0.91 (t, $J = 7.2$ Hz, 6H). **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 10.96 (br s, 1H), 8.33 (d, $J = 8.4$ Hz, 1H), 7.88–7.83 (m, 3H), 7.67–7.63 (m, 3H), 7.49–7.47 (m, 4H), 7.43–7.40 (m, 2H), 7.32–7.29 (m, 2H), 7.25–7.22 (m, 2H), 6.97 (d, $J = 2.2$ Hz, 2H), 6.92 (d, $J = 7.7$ Hz, 1H), 6.65–6.62 (m, 3H), 5.89 (s, 1H), 5.02 (d, $J = 7.5$ Hz, 1H), 4.01 (t, $J = 6.6$ Hz, 4H), 3.94 (d, $J = 7.5$ Hz, 1H), 3.73 (s, 2H), 2.51 (s, 3H), 1.84–1.78 (m, 4H), 1.49–1.36 (m, 8H), 0.95 (t, $J = 7.2$ Hz, 6H). **^{13}C NMR** (125.7 MHz, DMSO-d_6): δ 174.4, 173.2, 172.4, 165.5, 165.0, 159.8, 156.6, 151.5, 148.1, 142.9, 138.4, 138.2, 136.8, 135.8, 134.7, 134.1, 133.6, 130.0, 129.9, 128.4, 127.1, 127.0, 126.3, 126.2, 120.8, 119.1, 114.9, 111.7, 106.1, 104.0, 78.0, 68.1, 67.8, 56.5, 40.1, 28.4, 27.8, 23.6, 21.9, 14.0. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{55}H_{54}N_5O_9$ 928.3927, found 928.3934. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{55}H_{55}N_5NaO_9$ 952.3892, found 952.3877.

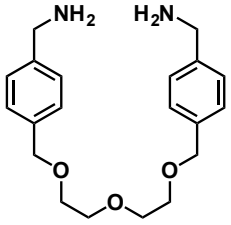
4,4'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(methylene))dibenzonitrile[Ref. 302,310], 163



4-(Bromomethyl)benzonitrile (5.00 g, 25.5 mmol), diethylene glycol (1.35 g, 12.8 mmol) and NaH (60% dispersion in oil, 1.28 g, 31.9 mmol) were added to a flame-dried flask containing dry THF (100 mL). The mixture was heated to 70 °C for 16 hours. The reaction was carefully quenched with H_2O and EtOH and the product was extracted into EtOAc, dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure, affording the desired dinitrile **163** as a white solid in sufficient purity for further synthetic modifications (4.27 g, 99%).

^{1}H NMR 499.9 MHz, $CDCl_3$): δ 7.62–7.61 (m, 4H), 7.45–7.44 (m, 4H), 4.62 (s, 4H), 3.72–3.67 (m, 8H). **^{13}C NMR** (75.5 MHz, $CDCl_3$): δ 144.0, 132.2, 127.8, 118.9, 111.3, 72.3, 70.7, 70.2.

((((Oxybis(ethane-2,1-diyl))bis(oxy))bis(methylene))bis(4,1-phenylene))dimethanamine[Ref. 302], 164

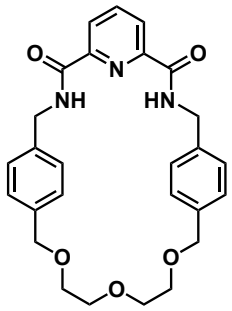


Dinitrile **163** (4.27 g, 12.7 mmol) was dissolved in dry THF (140 mL) under inert atmosphere and the solution was cooled on ice. BH_3·THF (4.92 g, 1 M, 57.3 mL, 57.3 mmol) was injected dropwise and the mixture was stirred at RT for two hours, followed by heating at 70 °C for 18 hours, at which point TLC analysis showed complete consumption of the dinitrile **163** (hexane:EtOAc, 1 : 1). The solution was allowed to cool to RT and the reaction was slowly quenched with *aq.* HCl (1 M, 50.0 mL). THF was removed under reduced pressure and the mixture was re-dissolved in CH_2Cl_2 and basified with *aq.* LiOH to pH 12. The organic fractions were dried over $MgSO_4$ and filtered. Solvent

was removed under reduced pressure and the diamine product **164** was obtained as a white solid, which used without further purification (3.94 g, 90%).

^{1}H NMR (500.1 MHz, $CDCl_3$): δ 7.33–7.26 (m, 8H), 4.55 (s, 4H), 3.85 (s, 4H), 3.69–3.62 (m, 8H), 1.58 (s, 4H).

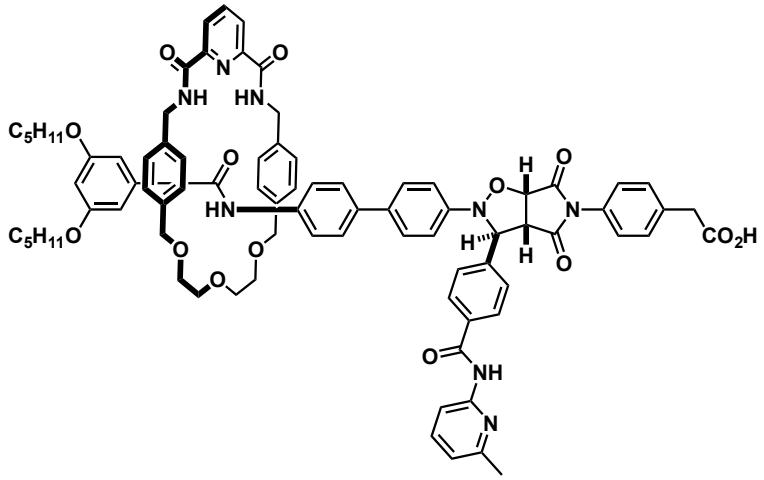
11,14,17-Trioxa-3,7-diaza-5(2,6)-pyridina-1,9(1,4)-dibenzenacyclooctadecaphane-4,6-dione, glycol macrocycle GM[Ref. 302, 310, 309]



Crude diamine **164** (1.00 g, 2.90 mmol) and pyridine-2, 6-dicarbonyl dichloride (0.59 g, 2.89 mmol) were dissolved in a portion of dry CH_2Cl_2 (50.0 mL) each and were slowly added dropwise to a solution of 4-pyridone (0.276 g, 2.90 mmol) and Et_3N (1.47 g, 2.00 mL, 14.5 mmol) in CH_2Cl_2 (250 mL) and THF (50.0 mL) under nitrogen. The mixture was stirred at RT for 5 days. The reaction mixture was concentrated under reduced pressure and the crude product was purified by flash silica chromatography ($CHCl_3$:EtOAc, 1 : 9), affording the target macrocycle **GM** as a white solid (0.523 g, 38%).

^{1}H NMR 400.1 MHz, $CDCl_3$): δ 8.30 ($J = 7.7$ Hz, 2H), 8.05–7.98 (m, 3H), 7.31–7.29 (m, 4H), 7.24–7.22 (m, 4H), 4.63 ($J = 5.7$ Hz, 4H), 4.58 (s, 4H), 3.74–3.68 (m, 8H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 163.3, 148.7, 139.5, 138.2, 136.9, 128.3, 127.6, 125.2, 72.9, 71.1, 69.8, 43.4.

Glycol macrocycle [2]rotaxane, GM-R

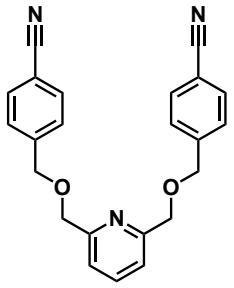


Linear component **L** (14.7 mg, 21.0 µmol) and maleimide S (4.86 mg, 21.0 µmol) were added to a solution of glycol macrocycle **GM** (100 mg, 21.0 µmol) in $CDCl_3$ (8.40 mL). The mixture was equilibrated in the freezer at −20 °C for 20 minutes and then stored at 4 °C. After two days, a further portion of Linear component **L** (14.7 mg, 21.0 µmol) and maleimide **S** (4.86 mg, 21.0 µmol) was added. The solution was analysed by **^{1}H NMR** spectroscopy 18 days after initial mixing, by which time the reaction was determined to be complete. Solvent was removed under a flow of air to afford the crude product, which contained **thread/GM [2]rotaxane** in a 1 : 1 ratio. The crude mixture was purified by reverse phase chromatography (20 g C18 column), eluting with MeCN:H_2O (3 : 1, 1.00 L), followed by MeCN:H_2O (4 : 1, 1.50 L). The **GM [2]Rotaxane** was isolated as a white solid (17.5 mg, 15%).

M.p.: 153–157 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 10.87 (s, 1H), 9.35–9.33 (m, 2H), 8.49 (d, $J = 7.9$ Hz, 2H), 8.31 (d, $J = 8.3$ Hz, 1H), 8.11 (t, $J = 7.9$ Hz, 1H), 7.95–7.93 (m, 2H), 7.71–7.68 (m, 2H), 7.60–7.59 (m, 2H), 7.52–7.50 (m, 4H), 7.42–7.40 (m, 2H), 7.35–7.33 (m, 2H), 7.29–7.27 (m, 2H), 6.92 (d, $J = 7.5$ Hz, 1H), 6.72–6.69 (m, 6H), 6.63–6.59 (m, 6H), 6.46 (t, $J = 2.1$ Hz, 1H), 5.89 (s, 1H), 5.07 (d, $J = 7.5$ Hz, 1H), 4.77–4.71 (m, 2H), 4.35–4.31 (m, 2H), 4.19–4.08 (m, 4H), 4.00 (d, $J = 7.5$ Hz, 1H), 3.82–3.69 (m, 6H), 3.64 (t, $J = 6.4$ Hz, 4H), 3.51–3.44 (m, 4H), 2.40 (s, 3H), 1.70–1.64 (m, 4H), 1.41–1.35 (m, 8H), 0.96–0.94 (m, 6H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 176.4, 174.8, 173.5, 166.5, 164.4, 164.1, 164.1, 159.5, 155.9, 151.6, 149.7, 149.6, 147.9, 142.5, 140.4, 138.8, 138.2, 136.8, 136.7, 135.9, 135.8, 135.7, 135.7, 135.7, 134.9, 133.9, 130.8, 130.0, 128.9, 128.9, 128.0, 127.6, 126.8, 126.4, 125.9, 125.4, 125.4, 120.4, 120.0, 115.2, 113.1, 106.9, 103.9, 77.7, 73.5, 73.5, 70.7, 68.9, 68.8, 68.8, 68.0, 57.3, 43.9, 43.9, 40.3, 29.1, 28.4, 22.6, 22.2, 14.2. **MS** (MALDI): *m/z* calculated

for $[M+H]^+$ $C_{82}H_{85}N_8O_{14}$ 1405.62, found 1405.60. **MS** (MALDI): *m/z* calculated for $[M+H]^+$ $C_{82}H_{85}N_8O_{14}$ 1405.62 (100%), 1406.62 (93%), 1407.63 (40%), 1408.63 (14%), found 1405.60 (100%), 1406.60 (76%), 1407.61 (44%), 1408.61 (20%).

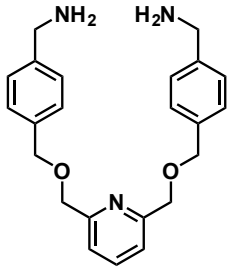
4,4'-(((Pyridine-2,6-diylbis(methylene))bis(oxy))bis(methylene))dibenzonitrile[Ref. 311, 303], 165



2,6-Pyridine dimethanol (1.77 g, 12.8 mmol) and 4-(bromomethyl)benzonitrile (5.00 g, 25.5 mmol) were added to a flame-dried flask and THF (80.0 mL) was added. NaH (60% dispersion in oil, 1.27 g, 31.9 mmol) was carefully added and the reaction mixture was heated to 70 °C for two days. The reaction was carefully quenched with H_2O (15.0 mL) and the product extracted into EtOAc, dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure, affording the desired dinitrile **165** as a white solid in sufficient purity for further synthetic modifications (4.69 g, 96%).

M.p.: 106.7–108.6 °C. **^{1}H NMR** (300.1 MHz, $CDCl_3$): δ 7.75 (t, $J = 7.7$ Hz, 1H), 7.66–7.62 (m, 4H), 7.51–7.47 (m, 4H), 7.40 (d, $J = 7.7$ Hz, 2H), 4.70 (s, 4H), 4.69 (s, 4H). **^{13}C NMR** (75.5 MHz, $CDCl_3$): δ 157.5, 143.6, 137.6, 132.4, 127.9, 120.4, 118.9, 111.6, 73.7, 72.0. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{23}H_{20}N_3O_2$ 370.1550, found 370.1545. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{23}H_{19}N_3NaO_2$ 392.1369, found 392.1363.

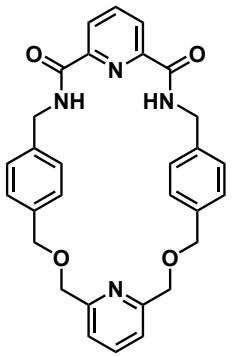
((((Pyridine-2,6-diylbis(methylene))bis(oxy))bis(methylene))bis(4,1-phenylene))dimethanamine[Ref. 303], 166



Dinitrile **165** (5.00 g, 13.5 mmol) was dissolved in dry THF (150 mL) under inert atmosphere and the solution was cooled on ice. BH_3·THF (5.23 g, 1 M, 60.9 mL, 60.9 mmol) was injected dropwise and the mixture was stirred at RT for two hours, followed by heating at 70 °C for 18 hours, at which point TLC analysis showed complete consumption of the dinitrile **165** (cyclohexane:EtOAc, 1 : 1). The solution was allowed to cool to RT and the reaction was slowly quenched with *aq.* HCl (1 M, 60.0 mL). THF was removed under reduced pressure and the mixture was re-dissolved in CH_2Cl_2 and basified with *aq.* KOH (2 M). The organic fractions were dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure and the diamine product **166** was obtained as a white solid, which used without further purification (4.62 g, 90%).

^{1}H NMR (400.1 MHz, $CDCl_3$): δ 7.71 (t, $J = 7.8$ Hz, 1H), 7.39 (d, $J = 7.8$ Hz, 2H), 7.37–7.34 (m, 4H), 7.32–7.29 (m, 4H), 4.66 (s, 4H), 4.63 (s, 4H), 3.86 (s, 4H), 1.57 (br s, 4H).

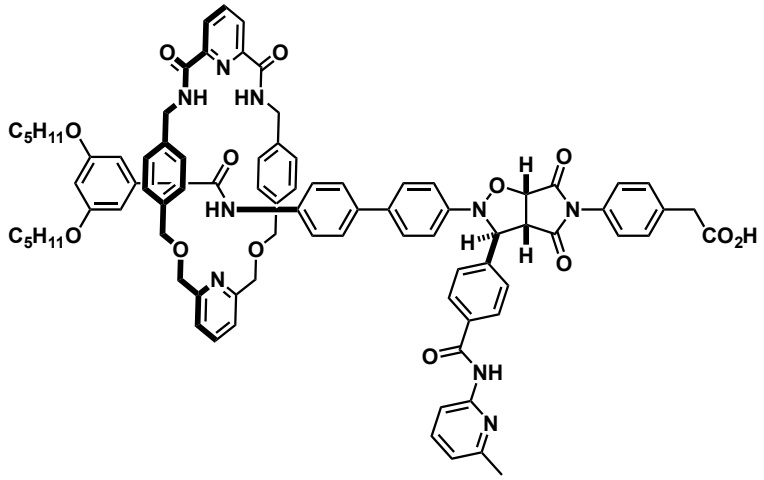
3,15-Dioxa-7,11-diaza-1,9(2,6)-dipyridina-5,13(1,4)-dibenzenacyclohexadecaphane-8,10-dione, pyridine macrocycle PM[Ref. 303, 311]



Crude diamine **166** (2.30 g, 6.09 mmol) and pyridine-2, 6-dicarbonyl dichloride (1.24 g, 6.08 mmol) were dissolved in a portion of dry CH_2Cl_2 (300 mL) each and were slowly added in portions (20.0 mL) at regular intervals (15 min) to a solution of Et_3N (5.80 g, 8.00 mL, 57.4 mmol) in CH_2Cl_2 (700 mL) under nitrogen. The mixture was stirred at RT for 13 days. The reaction mixture was concentrated under reduced pressure and the crude product was purified by flash silica chromatography (CH_2Cl_2:EtOAc, 6 : 1), affording the target macrocycle **PM** as a white solid (0.450 g, 11%).

M.p.: 219.5–221.6 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 8.39 (d, $J = 7.8$ Hz, 2H), 8.05 (t, $J = 7.8$ Hz, 1H), 7.93 (br t, $J = 5.0$ Hz, 2H), 7.70 (t, $J = 7.7$ Hz, 1H), 7.34–7.31 (m, 6H), 7.27–7.25 (m, 4H), 4.69 (d, $J = 6.1$ Hz, 4H), 4.62 (s, 4H), 4.52 (s, 4H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 163.5, 157.5, 148.8, 139.3, 137.5, 137.3, 137.3, 129.0, 127.8, 125.5, 121.2, 72.0, 71.9, 43.2. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{30}H_{29}N_4O_4$ 509.2183, found 509.2181. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{30}H_{28}N_4NaO_4$ 531.2003, found 531.1996.

Pyridine macrocycle [2]rotaxane, PM-R

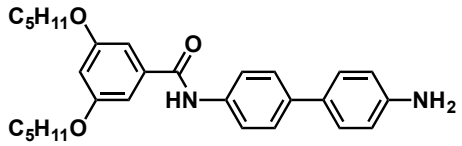


Macrocycle **PM** (104 mg, 0.204 mmol) and linear component **L** (23.8 mg, 0.0341 mmol) were dissolved in $CDCl_3$ (6.80 mL) and the mixture was equilibrated in the freezer at $-20\,^{\circ}C$ for one hour, followed by addition of stopper **S** (7.86 mg, 0.0340 mmol). After approximately 24 hours at $-20\,^{\circ}C$, the reaction mixture was stored at $4\,^{\circ}C$. After three days a second portion of **L** (23.8 mg, 0.0341 mmol) and maleimide **S** (7.86 mg, 0.0340 mmol) was added. The solution was analysed by **^{1}H NMR** spectroscopy nine days after initial mixing, by which time the reaction was determined to be complete. Solvent was removed under a flow of air to afford the crude product, which contained **thread/PM [2]rotaxane** in 1 : 2 ratio. The crude mixture was purified using by reverse phase chromatography (20 g C18 column), eluting with MeCN:H_2O (3 : 1, 1.00 L), followed by MeCN:H_2O (4 : 1, 1.00 L). The **PM [2]Rotaxane** was isolated as a 97 : 3 mixture of **PM [2]Rotaxane**:**thread** (15.6 mg, 16%).

^{1}H NMR (500.1 MHz, $CDCl_3$): δ 10.86 (br s, 1H), 9.53–9.50 (m, 2H), 9.16 (s, 1H), 8.49–8.47 (m, 2H), 8.33 (d, $J = 8.0$ Hz, 1H), 8.09 (t, $J = 7.8$ Hz, 1H), 7.95–7.92 (m, 2H), 7.76 (t, $J = 7.8$ Hz, 1H), 7.71 (t, $J = 8.0$ Hz, 1H), 7.49–7.47 (m, 4H), 7.36–7.32 (m, 4H), 7.28–7.27 (m, 2H), 7.25–7.24 (m, 2H), 7.20–7.18 (m, 2H), 6.92 (d, $J = 7.5$ Hz, 1H), 6.79–6.72 (m, 10H), 6.67–6.66 (m, 2H), 6.43 (t, $J = 2.1$ Hz, 1H), 5.85 (s, 1H), 5.04 (d, $J = 7.4$ Hz, 1H), 4.62–4.49 (m, 4H), 4.40–4.22 (m, 8H), 3.97 (d, $J = 7.4$ Hz, 1H), 3.74–3.66 (m, 2H), 3.47–3.40 (m, 4H), 2.41 (s, 3H), 1.58–1.50 (m, 4H), 1.33–1.19 (m, 8H), 0.93–0.89 (m, 6H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 176.4, 174.7, 173.4, 166.5, 165.2, 164.3, 164.3, 159.6, 157.2, 155.8, 151.6, 149.6, 149.5, 147.8, 142.5, 140.4, 138.8, 138.1, 137.8, 137.5, 137.4, 136.0, 135.7, 135.6, 135.3, 135.3, 134.9, 133.9, 130.8, 130.0, 129.2, 128.9, 128.0, 128.0, 127.6, 126.7, 126.3, 125.7, 125.4, 125.4, 121.9, 121.0, 120.0, 115.1, 113.1, 106.5, 105.4, 77.7, 73.3, 73.3, 71.4, 71.3,

68.9, 67.9, 57.3, 43.7, 43.7, 40.3, 29.0, 28.2, 22.6, 22.2, 14.2. **MS** (MALDI): *m/z* calculated for $[M+H]^+$ $C_{85}H_{84}N_9O_{13}$ 1438.62, found 1438.58. **MS** (MALDI): *m/z* calculated for $[M+H]^+$ $C_{85}H_{84}N_9O_{13}$ 1438.62 (100%), 1439.62 (96%), 1440.63 (43%), 1441.63 (16%), found 1438.58 (100%), 1439.59 (88%), 1440.59 (64%), 1441.60 (30%).

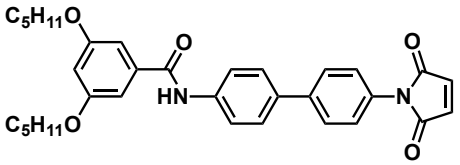
_N_-(4'-Amino-[1,1'-biphenyl]-4-yl)-3,5-bis(pentyloxy)benzamide, 170



Nitro compound **N** (0.661 g, 1.35 mmol) was dissolved in THF (50.0 mL) and Pd/C catalyst (10% wt, 0.100 g) was added. Sodium hypophosphite (0.642 g, 6.07 mmol) was dissolved in H_2O (20.0 mL) and the aqueous solution was added to THF mixture and heated to 40 °C. After three hours, a further portion of sodium hypophosphite (0.285 g, 2.69 mmol) in H_2O (5.00 mL) and Pd/C (10% wt, 50 mg) was added. After five hours overall, TLC analysis (EtOAc:hexane, 1 : 1) showed complete consumption of the starting material and the reaction mixture was filtered through celite, in order to remove the Pd catalyst. The aqueous THF solution was separated using EtOAc and brine, dried over $MgSO_4$ and the solvent was removed under reduced pressure. The amine product **170** was isolated as a yellow oil (0.55 g, 89%).

^{1}H NMR (499.9 MHz, $CDCl_3$): δ 7.78 (s, 1H), 7.67–7.64 (m, 2H), 7.55–7.53 (m, 2H), 7.43–7.40 (m, 2H), 6.97 (d, $J = 2.2$ Hz, 2H), 6.77–6.75 (m, 2H), 6.61 (t, $J = 2.2$ Hz, 1H), 4.00 (t, $J = 6.6$ Hz, 4H), 3.74 (s, 2H), 1.83–1.77 (m, 4H), 1.48–1.37 (m, 8H), 0.94 (t, $J = 7.2$ Hz, 6H). **^{13}C NMR** (125.7 MHz, $CDCl_3$): δ 165.6, 160.7, 145.9, 137.6, 137.2, 136.4, 131.0, 127.9, 127.0, 120.4, 115.6, 105.5, 104.7, 68.5, 29.0, 28.3, 22.6, 14.2. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{29}H_{37}N_2O_3$, 461.2799; found, 461.2792.

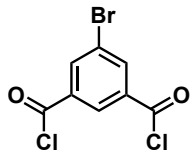
_N_-(4'-(2,5-Dioxo-2,5-dihydro-1H-pyrrol-1-yl)-[1,1'-biphenyl]-4-yl)-3,5-bis(pentyloxy)benzamide, linear component 167



Amine **170** (0.546 g, 1.19 mmol) and maleic anhydride (0.116 g, 1.19 mmol) were dissolved in dry THF (15.0 mL) and the solution was stirred at RT for 16 hours. The solution containing the crude uncyclised compound was concentrated down under a flow of compressed air and re-dissolved in AcOH (20.0 mL), transferred to a large glass microwave vial equipped with a snap-on cap (CEM) and stirred at RT for one hour, followed by heating to 120 °C using microwave irradiation for two hours. The resulting clear brown solution was diluted in CH_2Cl_2 and carefully washed with H_2O and saturated Na_2CO_3 solution. The organic fractions were washed with brine, dried over $MgSO_4$ and concentrated down under reduced pressure. The crude maleimide was purified by column chromatography (pet. ether 40–60 °C:EtOAc, 3 : 1) to give the target product **167** as a pale yellow solid (0.48 g, 29% over the two steps).

M.p.: 201.5–202.8 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 7.87 (s, 1H), 7.73–7.71 (m, 2H), 7.69–7.67 (m, 2H), 7.61–7.59 (m, 2H), 7.43–7.41 (m, 2H), 6.98 (d, $J = 2.2$ Hz, 2H), 6.89 (s, 2H), 6.62 (t, $J = 2.2$ Hz, 1H), 4.00 (t, $J = 6.6$ Hz, 4H), 1.81 (quintet, $J = 7.2$ Hz, 4H), 1.49–1.36 (m, 8H), 0.94 (t, $J = 7.2$ Hz, 6H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 169.7, 165.7, 160.7, 140.4, 137.7, 137.0, 136.4, 134.4, 130.3, 127.9, 127.7, 126.5, 120.5, 105.5, 104.7, 68.5, 29.0, 28.3, 22.6, 14.2. **HRMS** (ESI−): *m/z* calculated for $[M-H]^-$ $C_{33}H_{35}N_2O_5$, 539.2551; found, 539.2556.

5-Bromoisophthaloyl dichloride, 173

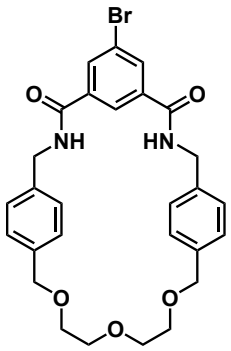


5-Bromoisophthalic acid (1.00 g, 4.08 mmol) was dissolved in $SOCl_2$ (20.0 mL) and DMF (8 drops) and the solution was heated to 100 °C for six hours, at which point TLC analysis (pet. ether 40–60 °C:EtOAc, 1 : 1) showed complete consumption of the starting material. The crude reaction mixture was concentrated under a flow of

compressed air and dried under high vacuum overnight, giving the crude activated acyl chloride **173** as an orange solid, in sufficient purity for immediate further conversion (quantitative conversion).

^{1}H NMR (500.1 MHz, $CDCl_3$): δ 8.76 (t, $J = 1.6$ Hz, 1H), 8.52 (d, $J = 1.6$ Hz, 2H).

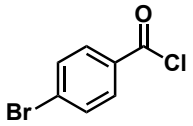
5^5-Bromo-11,14,17-trioxa-3,7-diaza-1,9(1,4),5(1,3)-tribenzenacyclooctadecaphane-4,6-dione, macrocycle 172



Glycol diamine **164** (3.94 g, 11.4 mmol) was dissolved in dry CH_2Cl_2 (100 mL) and dry THF (50.0 mL). In a separate flask, activated 5-bromoisophthaloyl dichloride **173** (3.22 g, 11.4 mmol) was dissolved in dry THF (150 mL). The diamine **164** and the di-acyl chloride **173** solutions were added in 10 portions over five hours to a solution of Et_3N (6.94 g, 68.6 mmol) in dry THF (1.70 L), and the mixture was stirred at RT under argon for five days. The mixture was concentrated under reduced pressure and the crude product was purified by flash column chromatography (CH_2Cl_2:EtOAc, $2:1 \rightarrow 1:1$) to give the desired macrocycle **172** as a white solid (1.28 g, 20%).

M.p.: 236.3–238.4 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$) δ 8.16 (s, 2H), 7.68 (s, 1H), 7.31–7.28 (m, 8H), 6.59 (s, 2H), 4.55 (d, $J = 5.3$ Hz, 4H), 4.48 (s, 4H), 3.67–3.60 (m, 8H). **^{13}C NMR** (125.8 MHz, $CDCl_3$) δ 165.3, 137.7, 137.1, 136.5, 134.2, 129.0, 128.7, 123.9, 122.1, 73.2, 70.6, 69.7, 44.6. **HRMS** (ESI+): *m/z* calculated for $[M+Na]^+$ $C_{28}H_{29}BrN_2NaO_5$, 575.1152; found, 575.1145.

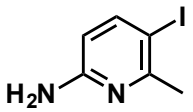
4-Bromobenzoyl chloride, 174



4-Bromobenzoic acid (2.50 g, 12.4 mmol) was added to dry PhMe (25.0 mL) and $SOCl_2$ (10.0 mL) and refluxed at 150 °C for three hours, at which point TLC analysis (CH_2Cl_2:EtOAc, 2 : 1) showed complete consumption of the starting material. The reaction mixture was concentrated under a flow of compressed air to give the crude compound **174** as a white solid, in sufficient purity for direct further reaction (quantitative conversion).

^{1}H NMR (500.1 MHz, $CDCl_3$): δ 7.99–7.97 (m, 2H), 7.68–7.66 (m, 2H).

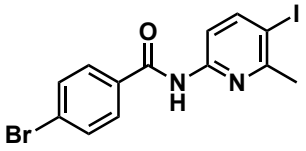
5-Iodo-6-methylpyridin-2-amine[Ref. 343], 175



2-Amino-6-picoline (5.00 g, 46.2 mmol) was dissolved in AcOH (50.0 mL) and H_2SO_4 (7.50 mL) and stirred at 30 °C. After 30 minutes, HIO_4 (1.33 g, 6.94 mmol) and I_2 (4.99 g, 19.7 mmol) were added and the mixture was stirred at 60 °C for 20 hours. AcOH was removed under reduced pressure and the remaining reaction mixture was diluted in H_2O. Aqueous KOH (5 M) was added to basify the mixture to pH 14, and the product was obtained by extraction with CH_2Cl_2. The organic fractions were washed with sodium thiosulfate and dried over $MgSO_4$. Concentration of the solution under reduced pressure afforded the crude product which was purified by recrystallisation from hot CH_2Cl_2, to give the desired iodinated product **175** as a pale brown crystalline solid (4.61 g, 43%).

M.p.: 101.8–103.4 °C. (lit.[343] 100 °C) **^{1}H NMR** (400.1 MHz, $CDCl_3$): δ 7.66 (dd, J = 8.5, 0.2 Hz, 1H), 6.10 (dq, J = 8.5, 0.6 Hz, 1H), 4.51 (s, 2H), 2.53 (s, 3H). **^{13}C NMR** (100.6 MHz, $CDCl_3$): δ 158.3, 157.7, 147.5, 108.0, 80.7, 28.6. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_6H_8N_2I$, 234.9732; found, 234.9734.

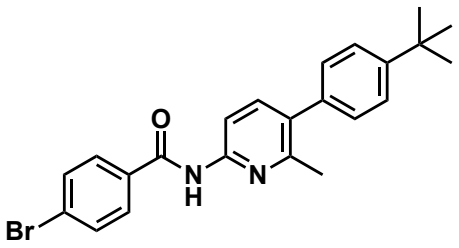
4-Bromo-*N*-(5-iodo-6-methylpyridin-2-yl)benzamide, 176



A solution of 2-amino-5-iodo-6-methylpyridine **175** (1.52 g, 6.49 mmol) in CH_2Cl_2 (20.0 mL) was added dropwise to a solution of 4-bromobenzoyl chloride **174** (1.50 g, 6.83 mmol) and Et_3N (1.39 g, 1.85 mL, 13.6 mmol) in CH_2Cl_2 (20.0 mL), cooled to 0 °C. The mixture was allowed to slowly warm up to RT and stirred for three days. The crude mixture was concentrated under reduced pressure and the crude product was re-dissolved in CH_2Cl_2 and precipitation of an insoluble salt was induced by addition of pet. ether 40–60 °C, which was removed by filtration under suction. Further precipitate formed during the filtration was isolated by another round of vacuum filtration, providing the target product **176** as a pale brown solid (1.05 g, 39%), which was used directly without further purification.

M.p.: 125.7–128.9 °C. **^{1}H NMR** (500.1 MHz, $CDCl_3$): δ 8.66 (s, 1H), 8.06 (d, J = 8.4 Hz, 1H), 7.94 (d, J = 8.4 Hz, 1H), 7.81–7.79 (m, 2H), 7.65–7.63 (m, 2H), 2.64 (s, 3H). **^{13}C NMR** (125.8 MHz, $CDCl_3$): δ 164.7, 158.7, 150.5, 148.5, 132.9, 132.2, 128.9, 127.4, 113.0, 89.2, 28.5. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{13}H_{11}BrIN_2O$, 416.9094; found, 416.9086.

4-Bromo-*N*-(5-(4-(*tert*-butyl)phenyl)-6-methylpyridin-2-yl)benzamide, 177

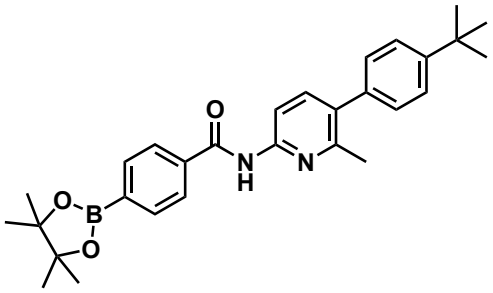


Iodo compound **176** (1.00 g, 2.40 mmol) and 4-*tert*-butylbenzeneboronic acid (0.470 g, 2.63 mmol) were added to a mixture of Na_2CO_3 (0.760 g, 7.17 mmol) in EtOH (9.00 mL), PhMe (9.00 mL) and H_2O (6.00 mL) and the mixture was degassed with argon for 30 minutes. $Pd(PPh_3)_4$ (0.138 g, 0.120 mmol) was added and the solution degassed for further 15 minutes, followed by heating to 60 °C for two days. The crude reaction mixture was diluted with CH_2Cl_2 and washed with brine. The organic fraction was dried over $MgSO_4$, filtered under gravity and concentrated under reduced

pressure. The crude product was purified by flash column chromatography (pet. ether 40–60 °C:CH_2Cl_2, 3 : 1 → 1 : 1) to give the product **177** as a white solid (0.72 g, 71%).

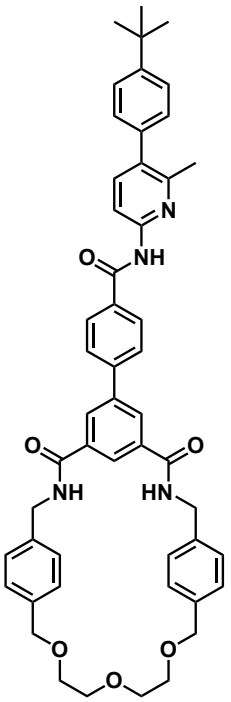
M.p.: 198.0–199.3 °C. **^{1}H NMR** (400.3 MHz, $CDCl_3$): δ 8.57 (s, 1H), 8.22 (d, J = 8.4 Hz, 1H), 7.84–7.80 (m, 2H), 7.66–7.61 (m, 3H), 7.47–7.44 (m, 2H), 7.28–7.25 (m, 2H), 2.45 (s, 3H), 1.37 (s, 9H). **^{13}C NMR** (100.7 MHz, $CDCl_3$): δ 164.7, 154.4, 150.5, 149.5, 140.1, 136.5, 133.6, 133.4, 132.2, 129.0, 129.0, 127.2, 125.5, 111.4, 34.8, 31.5, 23.1. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{23}H_{24}BrN_2O$, 423.1067; found, 423.1059.

N-(5-(4-(*Tert*-butyl)phenyl)-6-methylpyridin-2-yl)-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzamide, 178



Compound **177** (200 mg, 0.472 mmol), bis(pinacolato)diboron (220 mg, 0.866 mmol), KOAc (138 mg, 1.40 mmol) and $PdCl_2$(dppf)·CH_2Cl_2 (20.0 mg, 24.5 μmol) were added to dioxane (4.00 mL) and the mixture degassed for five minutes with argon and heated in a CEM microwave at 120 °C for three hours. The resulting solution was diluted with PhMe, washed with brine and dried over $MgSO_4$. The mixture was filtered through celite to remove the Pd catalyst and concentrated down under reduced pressure. The crude pinacol ester **178** was obtained as a yellow oil (quantitative conversion) that was used directly without purification.

***N*-(5-(4-(*Tert*-butyl)phenyl)-6-methylpyridin-2-yl)-4-(4,6-dioxo-11,14,17-trioxa-3,7-diaza-1,9(1,4),5(1,3)-tribenzenacyclooctadecaphane-5^5-yl)benzamide, 168**



Macrocycle **172** (0.130 g, 0.235 mmol), pinacol ester **178** (0.220 g, 4.68 mmol) and $CsCO_3$ (0.229 g, 0.703 mmol) were added to EtOH (9.00 mL), PhMe (9.00 mL) and H_2O (6.00 mL) and the mixture degassed with argon for 10 minutes. $Pd(PPh_3)_4$ (40.0 mg, 34.6 µmol) was added and the mixture degassed for further five minutes. The reaction mixture was heated to 90 °C for two days, diluted with CH_2Cl_2, and washed with H_2O. The organic fractions were dried over $MgSO_4$ and concentrated under reduced pressure. The crude product was purified by flash column chromatography (CH_2Cl_2:EtOAc, 1 : 2) to give the desired macrocycle **168** as a white solid (73.4 mg, 38%).

M.p.: > 340 °C (decomp.). **^{1}H NMR** (500.1 MHz, DMSO-d_6): δ 10.94 (s, 1H), 9.03 (t, J = 5.6 Hz, 2H), 8.28 (d, J = 1.4 Hz, 2H), 8.25–8.21 (m, 2H), 8.18 (t, J = 1.4 Hz, 1H), 8.15 (d, J = 8.4 Hz, 1H), 7.93–7.92 (m, 2H), 7.70 (d, J = 8.4 Hz, 1H), 7.51–7.49 (m, 2H), 7.36 (t, J = 3.6 Hz, 2H), 7.34–7.31 (m, 4H), 7.26–7.25 (m, 4H), 4.48 (d, J = 5.5 Hz, 4H), 4.40 (s, 4H), 3.51–3.46 (m, 8H), 2.46 (s, 3H), 1.34 (s, 9H). **^{13}C NMR** (125.8 MHz, DMSO-d_6): δ 166.0, 165.5, 153.6, 150.4, 149.6, 142.2, 139.4, 139.3, 138.6, 136.9, 136.3, 135.8, 133.5, 132.1, 129.0, 128.8, 128.5, 128.3, 128.1, 126.8, 125.5, 125.3, 112.2, 72.1, 69.9, 68.8, 43.0, 34.3, 31.2, 22.8. **HRMS** (ESI+): *m/z* calculated for $[M+H]^+$ $C_{51}H_{53}N_4O_7$, 817.3960; found, 817.3952.

REFERENCES

1. S. Milgram, *Psychol. Today* **1967**, *1*, 61–67.
2. H. Ebel, L. I. Mielsch, S. Bornholdt, *Phys. Rev. E* **2002**, *66*, 035103.
3. M. E. Newman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 404–409.
4. A. L. Barabési, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, *Physica A* **2002**, *311*, 590–614.
5. M. T. Gastner, M. E. Newman, *J. Stat. Mech.* **2006**, *1*, P01015.
6. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, *Comput. Networks* **2000**, *33*, 309–320.
7. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, *Nature* **2000**, *407*, 651–654.
8. P. Schuster, P. F. Stadler, *Complexity* **2002**, *8*, 34–42.
9. M. Mitchell, *Complexity: A Guided Tour*, Oxford University Press, New York, **2009**.
10. U.S.-Canada Power System Outage Task Force, Blackout in the United States and Canada: causes and recommendations, tech. rep., **2003**.
11. M. E. J. Newmann, *SIAM Rev.* **2003**, *45*, 167–256.
12. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. U. Hwang, *Phys. Rep.* **2006**, *424*, 175–308.
13. D. J. Watts, S. H. Strogatz, *Nature* **1998**, *393*, 440–442.
14. J. Gleick, *Chaos*, Minerva, **1987**.
15. M. Boots, A. Sasaki, *Proc. R. Soc. London B Biol. Sci.* **1999**, *266*, 1933–1938.

16. L. A. Amaral, A. Scala, M. Barthelemy, H. E. Stanley, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11149–111452.

17. J. Guare, Six degrees of separation: A play, New York, **1990**.

18. U. S. Bhalla, R. Iyengar, *Science* **1999**, *283*, 381–387.

19. A. Lazcano, *Chem. Biodivers.* **2008**, *5*, 1–15.

20. J. Gollihar, M. Levy, A. D. Ellington, *Science* **2014**, *343*, 259–260.

21. A. Lazcano, *Cold Spring Harb. Perspect. Biol.* **2010**, 1–17.

22. H. M. Leicester, *Development of Biochemical Concepts from Ancient to Modern Times*, Harvard University Press, Cambridge, **1974**.

23. C. Darwin, *On the Origin of the Species*, **1859**.

24. A. Pross, *J. Mol. Evol.* **2013**, *76*, 185–191.

25. C. Adami, C. Ofria, T. C. Collier, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4463–4468.

26. J.-M. Lehn, *Angew. Chem. Int. Ed.* **2013**, *52*, 2836–2850.

27. J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.

28. J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 964–967.

29. F. Crick, *Nature* **1970**, *227*, 561–563.

30. A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821–12844.

31. N. R. Pace, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 805–808.

32. N. Glansdorff, Y. Xu, B. Labedan, *Biol. Direct* **2008**, *3*, 1–35.

33. S. A. Tsokolov, *Astrobiology* **2009**, *9*, 401–412.

34. C. E. Cleland, C. F. Chyba, *Orig. Life Evol. Biosph.* **2002**, *32*, 387–393.

35. S. A. Benner, *Astrobiology* **2010**, *10*, 1021–1030.

36. D. L. Theobald, *Nature* **2010**, *465*, 219–222.

37. S. Tirard, *Orig. Life Evol. Biosph.* **2010**, *40*, 215–220.

38. *Origins of Life: the Central Concepts*, (Eds.: G. F. Joyce, D. W. Deamer, G. Fleischaker), Jones and Bartlett, Boston, **1994**, Foreword.

39. A. Pross, *J. Syst. Chem.* **2011**, *2*, 1.

40. A. J. Meyer, J. W. Ellefson, A. D. Ellington, *Acc. Chem. Res.* **2012**, *45*, 2097–2105.

41. A. Pross, R. Pascal, *Open Biol.* **2013**, *3*, 120190.

42. S. A. Kauffman, *Life* **2011**, *1*, 34–48.

43. K. Ruiz-Mirazo, C. Briones, A. de la Escosura, *Chem. Rev.* **2014**, *114*, 285–366.

44. C. Anastasi, F. F. Buchet, M. A. Crowe, A. L. Parkes, M. W. Powner, J. M. Smith, J. D. Sutherland, *Chem. Biodivers.* **2007**, *4*, 721–739.

45. J. W. Szostak, *J. Syst. Chem.* **2012**, *3*, 2.

46. M. Neveu, H.-J. Kim, S. A. Benner, *Astrobiology* **2013**, *13*, 391–403.

47. T. Czárán, B. Könnyu, E. Szathmáry, *J. Theor. Biol.* **2015**, *381*, 39–54.

48. L. E. Orgel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12503–12507.

49. E. Smith, H. J. Morowitz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13168–13173.

50. J. Peretó, *Chem. Soc. Rev.* **2012**, *41*, 5394–5403.

51. D. Segré, D. Lancet, *EMBO Rep.* **2000**, *1*, 217–222.

52. D. Segré, D. Ben-Eli, D. W. Deamer, D. Lancet, *Orig. Life Evol. Biosph.* **2001**, *31*, 119–145.

53. P. Razeto-Barry, *Orig. Life Evol. Biosph.* **2012**, *42*, 543–567.

54. A. Lazcano, *Orig. Life Evol. Biosph.* **2010**, *40*, 161–167.

55. M. Powner, J. Sutherland, *Phil. Trans. R. Soc. B* **2011**, *366*, 2870–2877.

56. G. M. Whitesides, *Science* **1999**, *284*, 89–92.

57. M. H. V. Regenmortel, *EMBO Rep.* **2004**, *5*, 1016–1020.

58. H. Kitano, *Science* **2002**, *295*, 1662–1664.

59. A. Friboulet, D. Thomas, *Biosens. Bioelectron.* **2005**, *20*, 2404–2407.

60. H.-Y. Chuang, M. Hofree, T. Ideker, *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 721–744.

61. J. J. Peyralans, S. Otto, *Curr. Opin. Chem. Biol.* **2009**, *13*, 705–713.

62. A. De la Escosura, C. Briones, K. Ruiz-Mirazo, *J. Theor. Biol.* **2015**, *381*, 11–22.

63. E. Mattia, S. Otto, *Nat. Nanotechnol.* **2015**, *10*, 111–119.

64. J. Stankiewicz, L. H. Eckardt, *Angew. Chem. Int. Ed.* **2006**, *45*, 342–344.

65. J. R. Nitschke, *Nature* **2009**, *462*, 736–738.

66. G. von Kiedrowski, S. Otto, P. Herdewijn, *J. Syst. Chem.* **2010**, *1*, 1.

67. V. Patzke, G. von Kiedrowski, *ARKIVOC* **2007**, *46*, 293–310.

68. Z. Dadon, N. Wagner, G. Ashkenasy, *Angew. Chem. Int. Ed.* **2008**, *47*, 6128–6136.

69. A. Vidonne, D. Philp, *Eur. J. Org. Chem.* **2009**, *2009*, 593–610.

70. A. J. Bissette, S. P. Fletcher, *Angew. Chem. Int. Ed.* **2013**, *52*, 12800–12826.

71. G. von Kiedrowski, *Bioorganic Chem. Front.* **1993**, *3*, 113–146.

72. E. Bigan, H.-P. Mattelaer, P. Herdewijn, *J. Mol. Evol.* **2016**, *82*, 93–109.

73. T. R. Cech, J. A. Steitz, *Cell* **2014**, *157*, 77–94.

74. X.-P. Song, M. Maiti, P. Herdewijn, *J. Syst. Chem.* **2011**, *2*, 3.

75. L. E. Orgel, R. Lohrmann, *Acc. Chem. Res.* **1974**, *7*, 368–377.

76. R. Lohrmann, L. E. Orgel, *J. Mol. Biol.* **1980**, *142*, 555–567.

77. T. Inoue, L. E. Orgel, *Science* **1983**, *219*, 859–862.

78. T. Inoue, G. F. Joyce, K. Grzeskowiak, L. E. Orgel, J. M. Brown, C. Reese, *J. Mol. Biol.* **1984**, *178*, 669–676.

79. G. von Kiedrowski, *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 932–935.

80. G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen, S. Jordan, *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 423–426.

81. A. Luther, R. Brandsch, G. von Kiedrowski, *Nature* **1998**, *396*, 245–248.

82. D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221–224.

83. D. Sievers, G. Von Kiedrowski, *Chem. Eur. J.* **1998**, *4*, 629–641.

84. J. Huck, D. Philp in *Supramolecular Chemistry: From Molecules to Nanomaterials*, John Wiley & Sons, Ltd., New York, **2012**, pp. 1415–1446.

85. N. Paul, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12733–12740.

86. J. Rogers, G. F. Joyce, *RNA* **2001**, *7*, 395–404.

87. D.-E. Kim, G. F. Joyce, *Chem. Biol.* **2004**, *11*, 1505–1512.

88. D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadri, *Nature* **1996**, *382*, 525–528.

89. E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science* **1991**, *254*, 539–544.

90. K. Severin, D. H. Lee, A. J. Kennan, M. R. Ghadiri, *Nature* **1997**, *389*, 706–709.

91. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776–779.

92. S. Yao, I. Ghosh, R. Zutshi, J. Chmielewski, *J. Am. Chem. Soc.* **1997**, *119*, 10559–10560.

93. S. Yao, I. Ghosh, R. Zutshi, J. Chmielewski, *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 478–481.

94. J. Y. Su, R. S. Hodges, C. M. Kay, *Biochemistry* **1994**, *33*, 15501–15510.

95. R. Fairman, H. G. Chao, L. Mueller, T. B. Lavoie, L. Shen, J. Novotny, G. R. Matsueda, *Protein Sci.* **1995**, *4*, 1457–1469.

96. R. Issac, J. Chmielewski, *J. Am. Chem. Soc.* **2002**, *124*, 6808–6809.

97. X. Li, J. Chmielewski, *J. Am. Chem. Soc.* **2003**, *125*, 11820–11821.

98. S. Yao, I. Ghosh, R. Zutshi, J. Chmielewski, *Nature* **1998**, *396*, 447–450.

99. D. H. Lee, K. Severin, Y. Yokobayashi, M. R. Ghadiri, *Nature* **1997**, *390*, 591–594.

100. K. Severin, D. H. Lee, J. A. Martinez, M. Vieth, M. R. Ghadiri, *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 126–128.

101. R. Plasson, H. Bersini, A. Commeyras, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16733–16738.

102. L. Plasson, D. K. Kondepudi, H. Bersini, A. Commeyras, K. Asakura, *Chirality* **2007**, *19*, 589–600.

103. D. G. Blackmond, *Chem. Eur. J.* **2007**, *13*, 3290–3295.

104. D. G. Blackmond, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, 1–18.

105. A. Saghatelian, Y. Yokobayashi, K. Soltani, M. R. Ghadiri, *Nature* **2001**, *409*, 797–801.

106. G. Ashkenasy, R. Jagasia, M. Yadav, M. R. Ghadiri, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10872–10877.

107. G. Ashkenasy, M. R. Ghadiri, *J. Am. Chem. Soc.* **2004**, *126*, 11140–11141.

108. Z. Dadon, M. Samiappan, E. Y. Safranchik, G. Ashkenasy, *Chem. Eur. J.* **2010**, *16*, 12096–12099.

109. T. Tjivikua, P. Ballester, J. Rebek, Jr., *J. Am. Chem. Soc.* **1990**, *112*, 1249–1250.

110. D. S. Kemp, K. S. Petrakis, *J. Org. Chem.* **1981**, *46*, 5140–5143.

111. F. M. Menger, A. V. Eliseev, N. A. Khanjin, *J. Am. Chem. Soc.* **1994**, *116*, 3613–3614.

112. D. N. Reinhoudt, D. M. Rudkevich, F. de Jong, *J. Am. Chem. Soc.* **1996**, *118*, 6880–6889.

113. V. Rotello, J.-I. Hong, J. J. Rebek, *J. Am. Chem. Soc.* **1991**, *113*, 9422–9423.

114. B. Wang, I. O. Sutherland, *Chem. Commun.* **1997**, 1495–1496.

115. M. Kindermann, I. Stahl, M. Reimold, W. M. Pankau, G. von Kiedrowski, *Angew. Chem. Int. Ed.* **2005**, *44*, 6750–6755.

116. F. Garcia-Tellado, S. Goswami, S.-K. Chang, S. J. Geib, A. D. Hamilton, *J. Am. Chem. Soc.* **1990**, *112*, 7393–7394.

117. R. J. Pearson, E. Kassianidis, D. Philp, *Tetrahedron Lett.* **2004**, *45*, 4777–4780.

118. R. J. Pearson, E. Kassianidis, A. M. Z. Slawin, D. Philp, *Org. Biomol. Chem.* **2004**, *2*, 3434–3441.

119. E. Kassianidis, R. J. Pearson, D. Philp, *Org. Lett.* **2005**, *7*, 3833–3836.

120. E. Kassianidis, R. J. Pearson, D. Philp, *Chem. Eur. J.* **2006**, *12*, 8798–8812.

121. R. J. Pearson, E. Kassianidis, A. M. Z. Slawin, D. Philp, *Chem. Eur. J.* **2006**, *12*, 6829–6840.

122. E. Kassianidis, D. Philp, *Chem. Commun.* **2006**, 4072–4074.

123. J. M. Quayle, A. M. Slawin, D. Philp, *Tetrahedron Lett.* **2002**, *43*, 7229–7233.

124. C. A. Booth, D. Phiip, *Tetrahedron Lett.* **1998**, *39*, 6987–6990.

125. V. C. Allen, D. Philp, N. Spencer, *Org. Lett.* **2001**, *3*, 777–780.

126. J. W. Sadownik, D. Philp, *Angew. Chem. Int. Ed.* **2008**, *47*, 9965–9970.

127. E. Kassianidis, D. Philp, *Angew. Chem. Int. Ed.* **2006**, *45*, 6344–6348.

128. R. J. Pieters, I. Huc, J. Rebek, *Angew. Chem. Int. Ed.* **1994**, *33*, 1579–1581.

129. R. J. Pieters, I. Huc, J. Rebek, *Tetrahedron* **1995**, *51*, 485–498.

130. E. Kassianidis, R. J. Pearson, E. A. Wood, D. Philp, *Faraday Discuss.* **2010**, *145*, 235–254.

131. V. C. Allen, C. C. Robertson, S. M. Turega, D. Philp, *Org. Lett.* **2010**, *12*, 1920–1923.

132. M. E. Belowich, J. F. Stoddart, *Chem. Soc. Rev.* **2012**, *41*, 2003–2024.

133. N. Giuseppone, *Acc. Chem. Res.* **2012**, *45*, 2178–2188.

134. G. Gasparini, M. Dal Molin, A. Lovato, L. J. Prins in *Supramolecular Chemistry: From Molecules to Nanomaterials*, John Wiley & Sons, Ltd., **2012**, pp. 1497–1526.

135. Y. Jin, C. Yu, R. J. Denman, W. Zhang, *Chem. Soc. Rev.* **2013**, *42*, 6634–6654.

136. J. D. Cheeseman, A. D. Corbett, J. L. Gleason, R. J. Kazlauskas, *Chem. Eur. J.* **2005**, *11*, 1708–1716.

137. P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652–3711.

138. I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106–2110.

139. K. S. Chichak, S. J. Cantrill, A. R. Pease, S.-H. Chiu, G. W. V. Cave, J. L. Atwood, J. F. Stoddart, *Science* **2004**, *304*, 1308–1312.

140. B. Brisig, J. K. M. Sanders, S. Otto, *Angew. Chem. Int. Ed.* **2003**, *42*, 1270–1273.

141. S. Otto, R. L. E. Furlan, J. K. M. Sanders, *J. Am. Chem. Soc.* **2000**, *122*, 12063–12064.

142. S. Otto, *Science* **2002**, *297*, 590–593.

143. S. M. Turega, C. Lorenz, J. W. Sadownik, D. Philp, *Chem. Commun.* **2008**, 4076–4078.

144. A. Buryak, K. Severin, *Angew. Chem. Int. Ed.* **2005**, *44*, 7935–7938.

145. Z. Dadon, M. Samiappan, N. Wagner, G. Ashkenasy, *Chem. Commun.* **2012**, *48*, 1419–1421.

146. A. Terfort, G. von Kiedrowski, *Angew. Chem. Int. Ed.* **1992**, *31*, 654–656.

147. S. Xu, N. Giuseppone, *J. Am. Chem. Soc.* **2008**, *130*, 1826–1827.

148. V. del Amo, A. M. Z. Slawin, D. Philp, *Org. Lett.* **2008**, *10*, 4589–4592.

149. J. Moore, N. Zimmerman, *Org. Lett.* **2000**, *2*, 915–918.

150. P. T. Corbett, S. Otto, J. K. M. Sanders, *Org. Lett.* **2004**, *6*, 1825–1827.

151. Z. Grote, R. Scopelliti, K. Severin, *Angew. Chem. Int. Ed.* **2003**, *42*, 3821–3825.

152. K. Severin, *Chem. Eur. J.* **2004**, *10*, 2565–2580.

153. P. T. Corbett, S. Otto, J. K. M. Sanders, *Chem. Eur. J* **2004**, *10*, 3139–3143.

154. R. F. Ludlow, S. Otto, *J. Am. Chem. Soc.* **2010**, *132*, 5984–5986.

155. B. De Bruin, P. Hauwert, J. N. H. Reek, *Angew. Chem. Int. Ed.* **2006**, *45*, 2660–2663.

156. P. T. Corbett, J. K. M. Sanders, S. Otto, *Angew. Chem. Int. Ed.* **2007**, *46*, 8858–8561.

157. V. del Amo, D. Philp, *Chem. Eur. J.* **2010**, *16*, 13304–13318.

158. J. Li, P. Nowak, S. Otto, *J. Am. Chem. Soc.* **2013**, *135*, 9222–9239.

159. J. D. Cheeseman, A. D. Corbett, R. Shu, J. Croteau, J. L. Gleason, R. J. Kazlauskas, *J. Am. Chem. Soc.* **2002**, *124*, 5692–5701.

160. A. D. Corbett, J. D. Cheeseman, R. J. Kazlauskas, J. L. Gleason, *Angew. Chem. Int. Ed.* **2004**, *43*, 2432–2436.

161. K. Osowska, O. Š. Miljanić, *Angew. Chem. Int. Ed.* **2011**, *50*, 8345–8349.

162. Q. Ji, O. S. Miljanić, *J. Org. Chem.* **2013**, *78*, 12710–12716.

163. W. L. Noorduin, T. Izumi, A. Millemaggi, M. Leeman, H. Meekes, W. J. P. Van Enckevort, R. M. Kellogg, B. Kaptein, E. Vlieg, D. G. Blackmond, *J. Am. Chem. Soc.* **2008**, *130*, 1158–1159.

164. J. M. A. Carnall, C. A. Waudby, A. M. Belenguer, M. C. A. Stuart, J. J.-P. Peyralans, S. Otto, *Science* **2010**, *327*, 1502–1506.

165. M. Colomb-Delsuc, E. Mattia, J. W. Sadownik, S. Otto, *Nat. Commun.* **2015**, *6*, 7427.

166. J. W. Sadownik, E. Mattia, P. Nowak, S. Otto, *Nat. Chem.* **2016**, *8*, 264–269.

167. B. Rubinov, N. Wagner, H. Rapaport, G. Ashkenasy, *Angew. Chem.* **2009**, *121*, 6811–6814.

168. B. Rubinov, N. Wagner, M. Matmor, O. Regev, N. Ashkenasy, G. Ashkenasy, *ACS Nano* **2012**, *6*, 7893–7901.

169. P. Vongvilai, M. Angelin, R. Larsson, O. Ramström, *Angew. Chem. Int. Ed.* **2007**, *46*, 948–950.

170. P. Vongvilai, R. Larsson, O. Ramström, *Adv. Synth. Catal.* **2008**, *350*, 448–452.

171. M. Angelin, P. Vongvilai, A. Fischer, O. Ramström, *Chem. Commun.* **2008**, *46*, 768–770.

172. M. Angelin, A. Fischer, O. Ramström, *J. Org. Chem.* **2008**, *73*, 3593–3595.

173. P. Vongvilai, O. Ramström, *J. Am. Chem. Soc.* **2009**, *131*, 14419–14425.

174. M. Sakulsombat, P. Vongvilai, O. Ramström, *Org. Biomol. Chem.* **2011**, *9*, 1112–1117.

175. Y. Zhang, L. Hu, O. Ramström, *Chem. Commun.* **2013**, *49*, 1805–1807.

176. M. Sakulsombat, Y. Zhang, O. Ramström, *Chem. Eur. J.* **2012**, *18*, 6129–6132.

177. M. Hutin, C. J. Cramer, L. Gagliardi, A. R. M. Shahi, G. Bernardinelli, R. Cerny, J. R. Nitschke, *J. Am. Chem. Soc.* **2007**, *129*, 8774–8780.

178. Y. Jin, C. Yu, R. J. Denman, W. Zhang, *Chem. Soc. Rev.* **2013**, *42*, 6634–6654.

179. Q. Ji, R. C. Lirag, O. Š. Miljanić, *Chem. Soc. Rev.* **2014**, *43*, 1873–1884.

180. J. W. Sadownik, D. Philp, *Org. Biomol. Chem.* **2015**, *13*, 10392–10401.

181. T. Kosikova, H. Mackenzie, D. Philp, *Chem. Eur. J.* **2016**, *22*, 1831–1839.

182. J.-M. Lehn, *Angew. Chem. Int. Ed.* **2015**, *54*, 2–16.

183. N. Paul, G. F. Joyce, *Curr. Opin. Chem. Biol.* **2004**, *8*, 634–639.

184. G. Ashkenasy, Z. Dadon, S. Alesebi, N. Wagner, N. Ashkenasy, *Isr. J. Chem.* **2011**, *51*, 106–117.

185. Z. Dadon, N. Wagner, S. Alasibi, M. Samiappan, R. Mukherjee, G. Ashkenasy, *Chem. Eur. J.* **2015**, *21*, 648–654.

186. J. Huck, PhD thesis, University of St Andrews, **2011**.

187. A. Vidonne, PhD thesis, University of St Andrews, **2009**.

188. E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, Sausalito, Califormia, **2006**.

189. D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution: Supplement*, Butterworths, London, **1972**.

190. C. Zhou, Y. Jin, J. R. Kenseth, M. Stella, K. R. Wehmeyer, W. R. Heineman, *J. Pharm. Sci.* **2005**, *94*, 576–589.

191. D. H. Ess, K. N. Houk, *J. Am. Chem. Soc.* **2008**, *130*, 10187–10198.

192. H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

193. *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*, (Eds.: A. Padwa, W. H. Pearson), John Wiley & Sons, Inc., NewYork, **2002**.

194. K. V. Gothelf, K. A. Jørgensen, *Chem. Rev.* **1998**, *98*, 863–909.

195. R. Sustmann, *Tetrahedron Lett.* **1971**, *29*, 2717–2720.

196. R. Sustmann, *Pure Appl. Chem.* **1974**, *40*, 569–593.

197. N. Coskun, A. Öztürk, *Tetrahedron* **2007**, *63*, 1402–1410.

198. B. Engels, M. Christl, *Angew. Chem. Int. Ed.* **2009**, *48*, 7968–7970.

199. C. C. Robertson, PhD thesis, University of St Andrews, **2011**.

200. M. I. Page, W. P. Jencks, *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 1678–1683.

201. M. I. Page, *Chem. Soc. Rev.* **1973**, *2*, 295–323.

202. A. J. Kirby, *Adv. Phys. Org. Chem.* **1980**, *17*, 183–278.

203. H. Mackenzie, MPhil Thesis, University of St Andrews, **2011**.

204. F. Dyson, *Nature* **2004**, *427*, 297.

205. J. Mayer, K. Khairy, J. Howard, *Am. J. Phys.* **2010**, *78*, 648–649.

206. I. R. Epstein, J. A. Pojman, O. Steinbock, *Chaos* **2006**, *16*, 037101.

207. J. W. S. Rayleigh, *Philos. Mag.* **1916**, *32*, 1–6.

208. M. Van Dyke, *An Album of Fluid Motion*, The Parabolic Press, Stanford, California, **1982**.

209. I. R. Epstein, K. Showalter, W. Virginia, *J. Chem. Phys.* **1996**, *100*, 13132–13147.

210. V. K. Vanag, I. R. Epstein, *Int. J. Dev. Biol.* **2009**, *53*, 673–681.

211. K. Showalter, I. R. Epstein, *Chaos* **2015**, *25*, 097613.

212. S. Kondo, T. Miura, *Science* **2010**, *329*, 1616–1620.

213. S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, E. Bonabeau, *Self-organisation in biological systems*, **2001**, p. 538.

214. P. Ball, *The Self-made Tapestry: Pattern Formation in Nature*, Oxford University Press Inc., New York, **2001**.

215. B. Hess, *Naturwissenschaften* **2000**, *87*, 199–211.

216. R. E. Liesegang, *Naturwissenschaftliche Wochenschrift* **1896**, *11*, 353–362.

217. S. C. Müller, J. Ross, *J. Phys. Chem. A* **2003**, *107*, 7997–8008.

218. A. M. Turing, *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **1952**, *641*, 37–72.

219. S. Kondo, T. Miura, T. Turing, *Science* **2010**, *329*, 1616–1620.

220. B. P. Belousov, *Sb. Ref. po Radiatsionni Meditsine* **1958**, 145.

221. A. M. Zhabotinsky, *Biofizika* **1964**, *9*, 306–311.

222. A. N. Zaikin, A. M. Zhabotinsky, *Nature* **1970**, *225*, 535–537.

223. R. Noyes, R. Field, E. Körös, *J. Am. Chem. Soc.* **1972**, *94*, 1394–1395.

224. R. J. Field, E. Körös, R. M. Noyes, *J. Am. Chem. Soc.* **1972**, *94*, 8649–8664.

225. I. R. Epstein, J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*, Oxford University Press, New York, **1998**.

226. A. J. Steele, M. Tinsley, K. Showalter, *Chaos* **2008**, *18*, 026108.

227. T. Bánsági, V. K. Vanag, I. R. Epstein, *Phys. Rev. E* **2012**, *86*, 045202.

228. I. R. Epstein, *Chem. Commun.* **2014**, *50*, 10758–67.

229. A. F. Taylor, M. R. Tinsley, K. Showalter, *Phys. Chem. Chem. Phys.* **2015**, *17*, 20047–20055.

230. E. Boga, G. Peintler, I. Nagypál, *J. Am. Chem. Soc.* **1990**, *112*, 151–153.

231. J. J. Tyson, J. D. Murray, *Development* **1989**, *106*, 421–426.

232. P. Foerster, S. C. Müller, B. Hess, *Development* **1990**, *16*, 11–16.

233. *Dictyostelids: Evolution, Genomics and Cell Biology*, (Eds.: M. Romeralo, S. Baldauf, R. Escalante), Springer, Berlin, **2013**.

234. G. J. Bauer, J. S. McCaskill, H. Otten, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7937–7941.

235. J. S. McCaskill, G. J. Bauer, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4191–4195.

236. A. Padirac, T. Fujii, A. Estévez-Torres, Y. Rondelez, *J. Am. Chem. Soc.* **2013**, *135*, 14586–14592.

237. A. S. Zadorin, Y. Rondelez, J. C. Galas, A. Estévez-Torres, *Phys. Rev. Lett.* **2015**, *114*, 1–5.

238. T. Fujii, Y. Rondelez, *ACS Nano* **2013**, *7*, 27–34.

239. J. H. Merkin, A. J. Poole, S. K. Scott, J. Masere, K. Showalter, *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 53–57.

240. A. Pross, *Orig. Life Evol. Biosph.* **2005**, *35*, 151–166.

241. I. Bottero, J. Huck, T. Kosikova, D. Philp, *J. Am. Chem. Soc.* **2016**, *138*, 6723–6726.

242. T. Kosikova, BSci Thesis, University of St Andrews, **2012**.

243. J. F. Stoddart, *Chem. Soc. Rev.* **2009**, *38*, 1802–1820.

244. G. Barin, R. S. Forgan, J. F. Stoddart, *Proc. R. Soc. A* **2012**, *468*, 2849–2880.

245. H. L. Frisch, E. Wasserman, *J. Am. Chem. Soc.* **1961**, *83*, 3789–3795.

246. D. A. Leigh, R. G. Pritchard, A. J. Stephens, *Nat. Chem.* **2014**, *6*, 978–982.

247. N. H. Evans, P. D. Beer, *Chem. Soc. Rev.* **2014**, *43*, 4658–4683.

248. G. Gil-Ramírez, D. A. Leigh, A. J. Stephens, *Angew. Chem. Int. Ed.* **2015**, *54*, 6110–6150.

249. M. Xue, Y. Yang, X. Chi, X. Yan, F. Huang, *Chem. Rev.* **2015**, *115*, 7398–7501.

250. C. O. Dietrich-Buchecker, J.-P. Sauvage, *Angew. Chem. Int. Ed.* **1989**, *28*, 189–192.

251. R. S. Forgan, J.-P. Sauvage, J. F. Stoddart, *Chem. Rev.* **2011**, *111*, 5434–5464.

252. J.-F. Ayme, J. E. Beves, C. J. Campbell, D. A. Leigh, *Chem. Soc. Rev.* **2013**, *42*, 1700–1712.

253. N. C. H. Lim, S. E. Jackson, *J. Phys. Condens. Matter* **2015**, *27*, 354101.

254. G. Schill, *Catenanes, Rotaxanes and Knots.* NY: Academic Press, New York, **1971**.

255. O. Safarowsky, B. Windisch, A. Mohry, F. Vögtle, *J. Prakt. Chem.* **2000**, *342*, 437–444.

256. A. Yerin, E. S. Wilks, G. P. Moss, A. Harada, *Pure Appl. Chem.* **2008**, *80*, 2041–2068.

257. E. Wasserman, *J. Am. Chem. Soc.* **1960**, *82*, 4433–4434.

258. I. T. Harrison, S. Harrison, *J. Am. Chem. Soc.* **1967**, *89*, 5723–5724.

259. G. Schill, A. Lüttringhaus, *Angew. Chem. Int. Ed. Engl.* **1964**, *3*, 546–547.

260. C. O. Dietrich-Buchecker, J.-P. Sauvage, J. P. Kintzinger, *Tetrahedron Lett.* **1983**, *24*, 5095–5098.

261. M. Cesario, C. O. Dietrich-Buchecker, J. Guilhem, C. Pascard, J.-P. Sauvage, *J. Chem. Soc. Chem. Commun.* **1985**, 244–247.

262. C. Wu, P. R. Lecavalier, Y. X. Shen, H. W. Gibson, *Chem. Mater.* **1991**, *3*, 569–572.

263. D. B. Amabilino, J. F. Stoddart, *Chem. Rev.* **1995**, *95*, 2725–2828.

264. S. Durot, F. Reviriego, J.-P. Sauvage, *Dalt. Trans.* **2010**, *39*, 10557–10570.

265. F. Aricò, J. D. Badjic, S. J. Cantrill, A. H. Flood, K. C. F. Leung, Y. Liu, J. F. Stoddart, *Top. Curr. Chem.* **2005**, *249*, 203–259.

266. K. D. Hänni, D. A. Leigh, *Chem. Soc. Rev.* **2010**, *39*, 1240–1251.

267. A. D. McNaught, A. Wilkinson, *The IUPAC Compendium of Chemical Terminology*, 2nd edn., Blackwell Scientific Publications, Oxford, **1997**.

268. M. C. T. Fyfe, J. F. Stoddart, *Acc. Chem. Res.* **1997**, *30*, 393–401.

269. P. R. Ashton, I. Baxter, M. C. T. Fyfe, F. M. Raymo, N. Spencer, J. F. Stoddart, A. J. P. White, D. J. Williams, *J. Am. Chem. Soc.* **1998**, *120*, 2297–2307.

270. M. C. T. Fyfe, F. M. Raymo, J. F. Stoddart, *Stimul. Concepts Chem.* **2005**, 211–220.

271. P. L. Anelli, P. R. Ashton, R. Ballardini, V. Balzani, M. Delgado, M. T. Gandolfi, T. T. Goodnow, A. E. Kaifer, D. Philp, *J. Am. Chem. Soc.* **1992**, *114*, 193–218.

272. F. G. Gatti, D. A. Leigh, S. A. Nepogodiev, A. M. Slawin, S. J. Teat, J. K. Wong, *J. Am. Chem. Soc.* **2001**, *123*, 5983–5989.

273. R. Jäger, S. Baumann, M. Fischer, O. Safarowsky, M. Nieger, F. Vögtle, *Liebigs Ann.* **1997**, 2269–2273.

274. A. H. Parham, B. B. Windisch, F. Vögtle, *Eur. J. Org. Chem.* **1999**, *1999*, 1233–1238.

275. P. R. Ashton, M. Belohradsky, D. Philp, J. F. Stoddart, *J. Chem. Soc. Chem. Commun.* **1993**, 1269–1274.

276. F. M. Raymo, K. N. Houk, J. F. Stoddart, *J. Am. Chem. Soc.* **1998**, *120*, 9318–9322.

277. V. Aucagne, J. Berné, J. D. Crowley, S. M. Goldup, K. D. Hänni, D. A. Leigh, P. J. Lusby, V. E. Ronaldson, A. M. Z. Slawin, A. Viterisi, D. B. Walker, *J. Am. Chem. Soc.* **2007**, *129*, 11950–11963.

278. J. D. Crowley, S. M. Goldup, A.-L. Lee, D. A. Leigh, R. T. McBurney, *Chem. Soc. Rev.* **2009**, *38*, 1530–1541.

279. S. M. Goldup, D. A. Leigh, P. R. McGonigal, V. E. Ronaldson, A. M. Z. Slawin, *J. Am. Chem. Soc.* **2010**, *132*, 315–320.

280. H. Lahlali, K. Jobe, M. Watkinson, S. M. Goldup, *Angew. Chem. Int. Ed.* **2011**, *50*, 4151–4155.

281. R. J. Bordoli, S. M. Goldup, *J. Am. Chem. Soc.* **2014**, *136*, 4817–4820.

282. V. Aucagne, K. D. Hänni, D. A. Leigh, P. J. Lusby, D. B. Walker, *J. Am. Chem. Soc.* **2006**, *128*, 2186–2187.

283. E. R. Kay, D. A. Leigh, *Top. Curr. Chem.* **2005**, *262*, 133–177.

284. P. L. Anelli, N. Spencer, J. F. Stoddart, *J. Am. Chem. Soc.* **1991**, *113*, 5131–5133.

285. R. Bissell, E. Cordova, A. E. Kaifer, J. F. Stoddart, *Nature* **1994**, *369*, 133–137.

286. E. R. Kay, D. A. Leigh, *Angew. Chem. Int. Ed.* **2015**, *54*, 10080–10088.

287. S. M. Goldup, *Nat. Nanotechnol.* **2015**, *10*, 488–489.

288. *Topics in Current Chemistry: Molecular Machines*, *Vol. 262*, (Ed.: T. R. Kelly), Springer, Germany, **2006**.

289. E. R. Kay, D. A. Leigh, *Pure Appl. Chem.* **2008**, *80*, 17–29.

290. J.-P. Sauvage, J.-P. Collin, S. Durot, J. Frey, V. Heitz, A. Sour, C. Tock, *C. R. Chim.* **2010**, *13*, 315–328.

291. M. von Delius, D. A. Leigh, *Chem. Soc. Rev.* **2011**, *40*, 3656–3676.

292. C. J. Bruns, J. F. Stoddart, *Acc. Chem. Res.* **2014**, *47*, 2186–2199.

293. S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, A. L. Nussbaumer, *Chem. Rev.* **2015**, *115*, 10081–10206.

294. R. F. Service, *Science* **2003**, *302*, 556–558.

295. A. Coskun, J. M. Spruell, G. Barin, W. R. Dichtel, A. H. Flood, Y. Y. Botros, J. F. Stoddart, *Chem. Soc. Rev.* **2012**, *41*, 4827–4859.

296. C. P. Collier, E. W. Wong, M. Belohradsky, F. M. Raymo, J. F. Stoddart, P. J. Kuekes, R. S. Williams, J. R. Heath, *Science* **1999**, *285*, 391–394.

297. Y. Luo, C. P. Collier, J. O. Jeppesen, K. A. Nielsen, E. DeIonno, G. Ho, J. Perkins, H. R. Tseng, T. Yamamoto, J. F. Stoddart, J. R. Heath, *ChemPhysChem* **2002**, *3*, 519–525.

298. J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin, E. DeIonno, Y. Luo, B. A. Sheriff, K. Xu, Y. S. Shin, H.-R. Tseng, J. F. Stoddart, J. R. Heath, *Nature* **2007**, *445*, 414–417.

299. B. Lewandowski, G. De Bo, J. W. Ward, M. Papmeyer, S. Kuschel, M. J. Aldegunde, P. M. E. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes, D. A. Leigh, *Science* **2013**, *339*, 189–193.

300. J. F. Stoddart, *Angew. Chem. Int. Ed.* **2014**, *53*, 11102–11104.

301. T. Kosikova, N. I. Hassan, D. B. Cordes, A. M. Z. Slawin, D. Philp, *J. Am. Chem. Soc.* **2015**, *137*, 16074–16083.

302. A. Vidonne, D. Philp, *Tetrahedron* **2008**, *64*, 8464–8475.

303. A. Vidonne, T. Kosikova, D. Philp, *Chem. Sci.* **2016**, *7*, 2592–2603.

304. N. I. Hassan, PhD thesis, University of St Andrews, **2011**.

305. P. Liu, W. Li, L. Liu, L. Wang, J. Ma, *J. Phys. Chem. A.* **2014**, *118*, 9032–9044.

306. L. Liu, Y. Liu, P. Liu, J. Wu, Y. Guan, X. Hu, C. Lin, Y. Yang, X. Sun, J. Ma, L. Wang, *Chem. Sci.* **2013**, *4*, 1701–1706.

307. N. I. Hassan, V. del Amo, E. Calder, D. Philp, *Org. Lett.* **2011**, *13*, 458–461.

308. S. Y. Hsueh, C. T. Kuo, T. W. Lu, C. C. Lai, Y. H. Liu, H. F. Hsu, S. M. Peng, C. H. Chen, S. H. Chiu, *Angew. Chem. Int. Ed.* **2010**, *49*, 9170–9173.

309. W.-C. Hung, L.-Y. Wang, C.-C. Lai, Y.-H. Liu, S.-M. Peng, S.-H. Chiu, *Tetrahedron Lett.* **2009**, *50*, 267–270.

310. Y.-L. Huang, W.-C. Hung, C.-C. Lai, Y.-H. Liu, S.-M. Peng, S.-H. Chiu, *Angew. Chem. Int. Ed.* **2007**, *46*, 6629–6633.

311. T. H. Ho, C. C. Lai, Y. H. Liu, S. M. Peng, S. H. Chiu, *Chem. Eur. J.* **2014**, *20*, 4563–4567.

312. D. M. D'Souza, D. A. Leigh, L. Mottier, K. M. Mullen, F. Paolucci, S. J. Teat, S. Zhang, *J. Am. Chem. Soc.* **2010**, *132*, 9465–9470.

313. I. A. O'Neil, A. J. Potter, J. M. Southern, A. Steiner, J. V. Barkley, *Chem. Commun.* **1998**, 2511–2512.

314. C. A. Hunter, D. H. Purvis, *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 792–795.

315. C. A. Schalley, *J. Phys. Org. Chem.* **2004**, *17*, 967–972.

316. D. A. Leigh, A. Murphy, J. P. Smart, A. M. Slawin, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 728–732.

317. P. R. Ashton, I. Baxter, S. J. Cantrill, M. C. T. Fyfe, P. T. Glink, J. F. Stoddart, A. J. P. White, D. J. Williams, *Angew. Chem. Int. Ed.* **1998**, *37*, 1294–1298.

318. S. J. Cantrill, G. J. Youn, J. F. Stoddart, D. J. Williams, *J. Org. Chem.* **2001**, *66*, 6857–6872.

319. S.-H. Chiu, S. J. Rowan, S. J. Cantrill, J. F. Stoddart, A. J. P. White, D. J. Williams, *Chem. Commun.* **2002**, 2948–2949.

320. R. Willst, F. Stadler, *Bull. Math. Biol.* **1998**, *60*, 1073–1098.

321. E. Szathmáry, I. Gladkih, *J. Theor. Biol.* **1989**, *138*, 55–58.

322. R. F. Heck, *Palladium Reagents in Organic Syntheses*, Academic Press Ltd., London, **1985**.

323. M. Karplus, *J. Chem. Phys.* **1959**, *30*, 11–15.

324. M. Karplus, *J. Am. Chem. Soc* **1963**, *85*, 2870–2871.

325. CrystalClear-SM Expert v2.0, Rigaku Americas, The Woodlands, Texas, USA, and Rigaku Corporation, Tokyo, Japan, 2010.

326. CrystalClear-SM Expert v2.1, Rigaku Americas, The Woodlands, Texas, USA, and Rigaku Corporation, Tokyo, Japan, 2013.

327. G. M. Sheldrick, *Acta Crystallogr. Sect. C.* **2015**, *71*, 3–8.

328. M. C. Burla, R. Caliandro, M. Camalli, B. Carrozzini, G. L. Cascarano, C. Giacovazzo, M. Mallamo, A. Mazzone, G. Polidori, R. Spagna, *J. Appl. Cryst.* **2012**, *45*, 357–361.

329. G. M. Sheldrick, *Acta Crystallogr. Sect. A.* **2008**, *64*, 112–122.

330. CrystalStructure v4.1. Rigaku Americas, The Woodlands, Texas, USA, and Rigaku Corporation, Tokyo, Japan, 2013.

331. M. C. Burla, R. Caliandro, M. Camalli, B. Carrozzini, G. L. Cascarano, L. De Caro, C. Giacovazzo, G. Polidori, R. Spagna, *J. Apl. Cryst.* **2005**, *38*, 381–388.

332. L. Palatinus, G. Chapuis, *J. Apl. Cryst.* **2007**, *40*, 786–790.

333. K. Rajesh, M. Somasundaram, R. Saiganesh, K. K. Balasubramanian, *J. Org. Chem.* **2007**, *72*, 5867–5869.

334. J. Richards, D. Philp, *Chem. Commun.* **2016**, *52*, 4995–4998.

335. B. J. Davie, C. Valant, J. M. White, P. M. Sexton, B. Capuano, A. Christopoulos, P. J. Scammells, *J. Med. Chem.* **2014**, *57*, 5405–5418.

336. M. A. Heuft, S. K. Collins, G. P. A. Yap, A. G. Fallis, *Org. Lett.* **2001**, *3*, 2883–2886.

337. A. Nierth, A. Y. Kobitski, G. Ulrich Nienhaus, A. Jäschke, *J. Am. Chem. Soc.* **2010**, *132*, 2646–2654.

338. S. Akiyama, K. Tajima, S. Nakatsuji, K. Nakashima, K. Abiru, M. Watanabe, *Bull. Chem. Soc. Jpn.* **1995**, *68*, 2043–2051.

339. S. C. Dickerman, A. M. Felix, L. B. Levy, *J. Org. Chem.* **1964**, *29*, 26–29.

340. F. P. Crisóstomo, T. Martín, R. Carrillo, *Angew. Chem. Int. Ed.* **2014**, *53*, 2181–2185.

341. Y.-X. Gao, L. Chang, H. Shi, B. Liang, K. Wongkhan, D. Chaiyaveij, A. S. Batsanov, T. B. Marder, C.-C. Li, Z. Yang, Y. Huang, *Adv. Synth. Catal.* **2010**, *352*, 1955–1966.

342. E. E. Nesterov, Z. Zhu, T. M. Swager, *J. Am. Chem. Soc.* **2005**, *127*, 10083–10088.

343. Z. Sun, S. Ahmed, L. W. McLaughlin, *J. Org. Chem.* **2006**, *71*, 2922–2925.

APPENDIX

A1. Example semi-automatic deconvolution script

```
-- Dinit.lua    semi-automatic deconvolution of a kinetic experiment

-- DEFINITIONS

HOME = HOME or "/Users/dpgroup/Desktop/TKT1dis/19F"      -- define your home directory
                                                  -- never forget the initial slash !
filename = "TK1recognitiondisabled.txt" --here we save the table of integrals

NumSpectra = 38 -- number of points for the kinetic study
NumRegions = 4  -- regions to deconvolute

P = 1           -- first spectrum to process
R = 1           -- first region to process
                        -- you can start from higher values, if you wish

F = {}          -- central frequencies
par = {}        -- parameter for the deconvolution
-- obtained from a preliminary deconvolution performed on the first spectrum
dx = 0.5        -- (constant) region to zoom in, in ppm units

local i = 1     -- progressive index, simplifies the editing of this script
-- for example, you can reorder the definitions below and they will still work

-- proton 1
F[i] = -101.6
par[i] = [[
Parameters for 1 peak
  frequency (Hz)     intensity      width (Hz)   Lorentzian %

      -51526.4         10.0000         10.0       100.0000
]]
i = i + 1

-- proton 2
F[i] = -109.745
par[i] = [[
Parameters for 1 peak
  frequency (Hz)     intensity      width (Hz)   Lorentzian %

      -56704.7         10.0000         10.0       100.0000
]]
i = i + 1

-- proton 3
F[i] = -115.86
par[i] = [[
Parameters for 1 peak
  frequency (Hz)     intensity      width (Hz)   Lorentzian %

      -51526.4         10.0000         10.0       100.0000
]]
i = i + 1

-- proton 4
```

```
F[i] = -120.35
par[i] = [[
Parameters for 1 peak
  frequency (Hz)     intensity      width (Hz)   Lorentzian %

      -56704.7         10.0000         10.0       100.0000
]]
i = i + 1

-- end of definitions ----------------------------------------

-- ************->    HERE WE GO:    <-***************

io.output(HOME..filename)       -- create/open the file were the results will be stored
io.write("point")       -- header
for i = 1,NumRegions do
        io.write("\tppm\tarea")
end
io.write("\n")

spectral = getf("x")                                    -- read some experimental parameters
conversion = 1.0 / spectral.MHz                  -- useful to convert from Hz to ppm
spectral = getf("y")
step = spectral.width / spectral.size
Y = spectral.start + step * (spectral.size -0.1 -P+1) -- position of the first row

io.write( string.format("\n\t%02d\t", P ) )              -- report the experiment no.
mark('h', Y )                 -- choose a row
extract()                     -- extract the coresponding 1D spectrum
delint()                        -- we need to normalize the intensities
region( -105, -125.0 )        -- region containing protons no. 1 and 9
press 'i'                       -- first integral, automatically set to 1
intreg( 1, 300 )              -- we set it to 300 to have manageable numbers (>1 and <300)

region( F[R] + dx, F[R] - dx )
press "z"
adj()
print("select the signal, then run fnext")
selection = true
```

A2. Example SimFit files employed in kinetic fitting

A2.1 SimFit input file for a simple bimolecular reaction

```
*================================================================================
* Simple bimolecular cycloaddition fit for T1 recognition disabled
* NITRONE = Amidopyridine + F-phenyl
* MAL     = Phenylacetic acid maleimide M1
*================================================================================
DIM ( 2 )

* Bimolecular routes to TRANS
REACTION ( NITRONE + MAL --> TRANS )  CONSTANT ( 1, 1E-4, 1, 1, 1000 )
REACTION ( NITRONE + MAL --> CIS   )  CONSTANT ( 2, 1E-4, 2, 1, 1000 )

REACTION (COMPILE)
REACTION (SHOW)
CONSTANT (SHOW)

DEFINE   (1, TRANS   , P, 1) SCALE (3,1)
DEFINE   (2, CIS     , P, 3) SCALE (3,1)
DEFINE   (3, NITRONE , P, 5) SCALE (3,1)

SELECT   ( NITRONE, CIS, TRANS )

READ     ( T1C )
REACTION ( DOC  )
CONSTANT ( DOC  )

TIME     (SEC)
WIN      (0, 60000, 20000, 200, 0, 5E-3, 1E-3, 1E-4)

ASSIGN   (OBS,  CIS     = CIS     )
ASSIGN   (OBS,  TRANS   = TRANS   )
ASSIGN   (OBS,  NITRONE = NITRONE )

ASSIGN   (SPEC, TRANS   = TRANS   )
ASSIGN   (SPEC, MAL     = #5e-3   )
ASSIGN   (SPEC, CIS     = CIS     )
ASSIGN   (SPEC, NITRONE = NITRONE )
CHOOSE   (EXP1)

INTEG    (STIFF, 1E-8, 200, 0.05, 200, 100)
PLOT     (OBS, RES)

OPAR     (1E16)
SIMPLEX  (PLOT)
SIMPLEX  (PLOT)
SIMPLEX  (PLOT)
SIMPLEX  (PLOT)
NEWTON   (PLOT)
PLOT     (FILE)
PLOT     (OBS, RES)
```

A2.2 SimFit input file for a recognition-enabled reaction

```
*================================================================================
* Kinetics for recognition-enabled formation of replicator T1
*================================================================================
* Ka for individual binding event is 830 M-1 (Koff = 1.2408E+6)
* A = Nitrone Nf, B = Maleimide M1
*================================================================================
DIM (2)

* Bimolecular routes to TRANS
REACTION (A + B          --> TRANS      )     CONSTANT ( 1, 1.1194E-4, 0)

* Bimolecular routes to CIS
REACTION (A + B          --> CIS        )     CONSTANT ( 2, 4.3089E-5, 0)

* Formation of binary complexes including product duplexes
* Only [TRANS*TRANS] is stable beyond one Pyr*COOH association
REACTION (A + TRANS     ==> ATRANS    ) CONSTANT ( 3, 1E9, 0) CONSTANT ( 4, 1.2408E6, 0)
REACTION (B + TRANS     ==> BTRANS    ) CONSTANT ( 5, 1E9, 0) CONSTANT ( 6, 1.2408E6, 0)
REACTION (A + BTRANS    ==> ABTRANS   ) CONSTANT ( 7, 1E9, 0) CONSTANT ( 8, 1.2408E6, 0)
REACTION (B + ATRANS    ==> ABTRANS   ) CONSTANT ( 9, 1E9, 0) CONSTANT (10, 1.2408E6, 0)
REACTION (A + CIS       ==> ACIS      ) CONSTANT (11, 1E9, 0) CONSTANT (12, 1.2408E6, 0)
REACTION (B + CIS       ==> BCIS      ) CONSTANT (13, 1E9, 0) CONSTANT (14, 1.2408E6, 0)
REACTION (A + BCIS      ==> ABCIS     ) CONSTANT (15, 1E9, 0) CONSTANT (16, 1.2408E6, 0)
REACTION (B + ACIS      ==> ABCIS     ) CONSTANT (17, 1E9, 0) CONSTANT (18, 1.2408E6, 0)
REACTION (TRANS + TRANS ==> TRANSTRANS) CONSTANT (19, 1E9, 0)
                                        CONSTANT (20, 1.000E2, 1, 1, 1000)
REACTION (CIS  + CIS    ==> CISCIS    ) CONSTANT (21, 1E9, 0) CONSTANT (22, 1.2408E6, 0)
REACTION (TRANS + CIS   ==> TRANSCIS  ) CONSTANT (23, 1E9, 0) CONSTANT (24, 1.2408E6, 0)
REACTION (A + B         ==> AB        ) CONSTANT (25, 1E9, 0) CONSTANT (26, 1.2408E6, 0)

* Ternary complex reaction
REACTION (ABTRANS      --> TRANSTRANS            )  CONSTANT (27, 1E-2, 2, 1, 1000)

*Bimolecular Reactions of Complexes
REACTION (AB + A       --> TRANS + A             )  CONSTANT (28, 1.1194E-4, 0    )
REACTION (AB + B       --> TRANS + B             )  CONSTANT (29, 1.1194E-4, 0    )
REACTION (AB + A       --> CIS + A               )  CONSTANT (30, 4.3089E-5, 0    )
REACTION (AB + B       --> CIS + B               )  CONSTANT (31, 4.3089E-5, 0    )

REACTION (AB + ACIS    --> TRANS + A + CIS       )  CONSTANT (32, 1.1194E-4, 0    )
REACTION (AB + BCIS    --> TRANS + B + CIS       )  CONSTANT (33, 1.1194E-4, 0    )
REACTION (AB + ATRANS  --> TRANS + A + TRANS     )  CONSTANT (34, 1.1194E-4, 0    )
REACTION (AB + BTRANS  --> TRANS + B + TRANS     )  CONSTANT (35, 1.1194E-4, 0    )

REACTION (AB + ACIS    --> CIS  + A + CIS        )  CONSTANT (36, 4.3089E-5, 0    )
REACTION (AB + BCIS    --> CIS  + B + CIS        )  CONSTANT (37, 4.3089E-5, 0    )
REACTION (AB + ATRANS  --> CIS  + A + TRANS      )  CONSTANT (38, 4.3089E-5, 0    )
REACTION (AB + BTRANS  --> CIS  + B + TRANS      )  CONSTANT (39, 4.3089E-5, 0    )

REACTION (AB + ABCIS   --> CIS  + A + B + CIS    )  CONSTANT (40, 4.3089E-5, 0    )
REACTION (AB + ABCIS   --> TRANS + A + B + CIS   )  CONSTANT (41, 1.1194E-4, 0    )
REACTION (AB + ABTRANS --> CIS  + A + B + TRANS  )  CONSTANT (42, 4.3089E-5, 0    )
REACTION (AB + ABTRANS --> TRANS + A + B + TRANS )  CONSTANT (43, 1.1194E-4, 0    )

REACTION (       2 AB  --> TRANS + A + B         )  CONSTANT (44, 1.1194E-4, 0    )
```

```
REACTION (       2 AB  --> CIS  + A + B          )  CONSTANT (45, 4.3089E-5, 0    )

*REACTION (        AB  --> CIS                   )  CONSTANT (46, 1E-8, 3, 1, 1000)

REACTION ( COMPILE )
REACTION ( SHOW    )
CONSTANT ( SHOW    )

DEFINE   ( 1, TRANS , P, 1) SCALE (3,1)
DEFINE   ( 2, A     , E, 3) SCALE (3,1)
DEFINE   ( 3, CIS   , P, 5) SCALE (3,1)

SELECT   ( A, CIS, TRANS )
READ     ( T1exp         )
REACTION ( DOC           )
CONSTANT ( DOC           )
TIME     ( SEC           )

WIN      ( 0, 80000, 15000, 200, 0, 5.2e-3, 1e-3, 3e-4)

ASSIGN   ( OBS,  TRANS = TRANS + ATRANS + BTRANS + ABTRANS + 2 TRANSTRANS + TRANSCIS )
ASSIGN   ( OBS,  A     = A     + AB     + ATRANS + ABTRANS + ACIS         + ABCIS    )
ASSIGN   ( OBS,  CIS   = CIS   + ACIS   + ABCIS  + BCIS + TRANSCIS + 2 CISCIS        )

ASSIGN   ( SPEC, A = A         )
ASSIGN   ( SPEC, B = #5E-3     )
ASSIGN   ( SPEC, TRANS = TRANS )
ASSIGN   ( SPEC, CIS = CIS     )

CHOOSE   ( EXP1, EXP2 )
OPAR     ( 1E16       )
INTEG    (STIFF, 1E-10, 50, 0.025, 100, 50)
PLOT     (OBS, RES    )

* Optimise rate constant using simplex
SIMPLEX  ( PLOT   )
NEWTON   ( PLOT   )
PLOT     ( FILE   )
PLOT     (OBS, RES)
```

A2.3 SimFit simulation file for a recognition-enabled reaction

```
*================================================================================
* T2 kinetic simulations
*================================================================================
MODE ( ISOSIM )

* Bimolecular reactions
REACTION (A + B          --> TRANS       , 1.7940E-4         )
REACTION (A + B          --> CIS         , 7.0364E-5         )

* Formation of binary complexes including product duplexes
REACTION (A + TRANS      ==> ATRANS     , 1.00E+9, 3.012E+5  )
REACTION (B + TRANS      ==> BTRANS     , 1.00E+9, 3.012E+5  )
REACTION (A + BTRANS     ==> ABTRANS    , 1.00E+9, 3.012E+5  )
REACTION (B + ATRANS     ==> ABTRANS    , 1.00E+9, 3.012E+5  )
REACTION (A + CIS        ==> ACIS       , 1.00E+9, 3.012E+5  )
REACTION (B + CIS        ==> BCIS       , 1.00E+9, 3.012E+5  )
REACTION (A + BCIS       ==> ABCIS      , 1.00E+9, 3.012E+5  )
REACTION (B + ACIS       ==> ABCIS      , 1.00E+9, 3.012E+5  )
REACTION (TRANS + TRANS  ==> TRANSTRANS , 1.00E+9, 1.4577E+1 )
REACTION (CIS  + CIS     ==> CISCIS     , 1.00E+9, 3.012E+5  )
REACTION (TRANS + CIS    ==> TRANSCIS   , 1.00E+9, 3.012E+5  )
REACTION (A + B          ==> AB         , 1.00E+9, 3.012E+5  )

* Ternary complex reaction
REACTION (ABTRANS      --> TRANSTRANS   , 1.6990E-3          )

*Bimolecular Reactions of Complexes
REACTION (AB + A       --> TRANS + A             , 1.7940E-4 )
REACTION (AB + B       --> TRANS + B             , 1.7940E-4 )
REACTION (AB + A       --> CIS + A               , 7.0364E-5 )
REACTION (AB + B       --> CIS + B               , 7.0364E-5 )
REACTION (AB + ACIS    --> TRANS + A + CIS       , 1.7940E-4 )
REACTION (AB + BCIS    --> TRANS + B + CIS       , 1.7940E-4 )
REACTION (AB + ATRANS  --> TRANS + A + TRANS     , 1.7940E-4 )
REACTION (AB + BTRANS  --> TRANS + B + TRANS     , 1.7940E-4 )
REACTION (AB + ACIS    --> CIS  + A + CIS        , 7.0364E-5 )
REACTION (AB + BCIS    --> CIS  + B + CIS        , 7.0364E-5 )
REACTION (AB + ATRANS  --> CIS  + A + TRANS      , 7.0364E-5 )
REACTION (AB + BTRANS  --> CIS  + B + TRANS      , 7.0364E-5 )
REACTION (AB + ABCIS   --> CIS  + A + B + CIS    , 7.0364E-5 )
REACTION (AB + ABCIS   --> TRANS + A + B + CIS   , 1.7940E-4 )
REACTION (AB + ABTRANS --> CIS  + A + B + TRANS  , 7.0364E-5 )
REACTION (AB + ABTRANS --> TRANS + A + B + TRANS , 1.7940E-4 )
REACTION (       2 AB  --> TRANS + A + B         , 1.7940E-4 )
REACTION (       2 AB  --> CIS  + A + B          , 7.0364E-5 )
REACTION (         AB  --> CIS                   , 2.3067E-6 )

REACTION ( COMPILE )
REACTION ( SHOW    )
CONSTANT ( SHOW    )

INIT   ( A     , 0.005, 1)
INIT   ( B     , 0.005, 2)
INIT   ( TRANS , 0.002, 3)
```

```
INTEG    (STIFF, 1E-10, 50, 0.025, 100, 50)
NUMPLOT  ( 100  )
TIME     ( SEC  )
WIN      ( 0, 80000, 15000, 200, 0, 5.2e-3, 1e-3, 3e-4)
PLOT     ( FILE )
PLOT
```

A3. Component concentrations for dynamic covalent libraries

A3.1 Control exchange pool

Table 9.6 Raw data describing the distribution of the exchange pool components in the control unin-structed experiment after two (blue) and seven days (red). Concentrations were quantitatively determined by $^{19}F\{^{1}H\}$ NMR spectroscopy relative to 1-bromo-2-fluoro-4-nitrobenzene.

Exchange Pool Components (mM)		A	B	C	D
		2.8 (2.8)	2.4 (2.3)	0.8 (0.7)	1.2 (1.2)
W	1.3 (1.3)	2.0 (2.1)	2.3 (2.3)	2.6 (2.7)	2.5 (2.6)
X	5.4 (5.4)	1.2 (1.2)	1.3 (1.3)	1.3 (1.3)	1.4 (1.4)
Y	2.9 (2.9)	1.7 (1.7)	1.9 (1.9)	1.9 (1.9)	1.9 (2.0)
Z	0.0 (0.1)	2.4 (2.3)	2.2 (2.2)	3.4 (3.3)	2.9 (2.9)

A3.2 M1–M2 dynamic covalent libraries

Table 9.7 Raw data describing the distribution of the exchange and product pool components in the **M1**–**M2** experiment after two (blue) and seven (red) days for **(a)** native, **(b)** 10 mol% of *trans*-**T1** **(c)** 10 mol% of *trans*-**T2** and **(d)** 10 mol% of *trans*-**T1** and *trans*-**T2**. Concentrations were quantitatively determined by $^{19}F\{^{1}H\}$ NMR spectroscopy relative to 1-bromo-2-fluoro-4-nitrobenzene. *Trans* product concentrations are corrected for the template added. **Cis* products are not subtracted.

(a) No template

Exchange Pool Components (mM)		A	B	C	D
		2.8* (3.2)*	2.5* (3.2)*	1.0* (1.5)*	1.4* (2.0)*
W	1.5 (1.4)	1.7 (1.2)	2.4 (2.4)	2.9 (3.0)	2.7 (2.8)
X	5.8 (5.5)	0.9 (0.7)	1.3 (1.4)	1.4 (1.5)	1.4 (1.5)
Y	3.2 (3.0)	1.5 (1.1)	2.0 (1.9)	2.1 (2.2)	2.1 (2.1)
Z	0.0 (0.0)	1.1 (0.2)	1.3 (0.4)	1.9 (0.6)	1.7 (0.4)
***Trans* Products**	**M1**	0.9 (1.7)	0.2 (0.3)	0.2 (0.4)	0.3 (0.4)
	M2	1.1 (1.9)	0.3 (0.5)	0.4 (0.7)	0.4 (0.7)

(b) + *trans*-**T1**

Exchange Pool Components (mM)		A	B	C	D
		2.0* (2.6)*	2.4* (3.1)*	1.0* (1.6)*	1.3* (1.9)*
W	1.7 (1.6)	1.5 (1.1)	2.5 (2.6)	3.0 (3.2)	2.9 (3.0)
X	5.9 (5.7)	0.9 (0.6)	1.4 (1.4)	1.5 (1.6)	1.5 (1.5)
Y	3.3 (3.1)	1.3 (0.9)	2.1 (2.1)	2.2 (2.3)	2.2 (2.3)
Z	0.0 (0.0)	0.8 (0.1)	1.1 (0.3)	1.7 (0.4)	1.4 (0.4)
***Trans* Products**	**M1**	1.9 (2.5)	0.2 (0.3)	0.2 (0.4)	0.2 (0.4)
	M2	1.6 (2.2)	0.3 (0.4)	0.4 (0.6)	0.4 (0.6)

(c) + *trans*-**T2**

Exchange Pool Components (mM)		A	B	C	D
		2.3* (2.9)*	2.3* (3.2)*	0.7* (1.6)*	1.1* (2.0)*
W	1.6 (1.4)	1.7 (1.2)	2.5 (2.4)	3.0 (3.0)	2.9 (2.9)
X	5.8 (5.4)	1.0 (0.7)	1.4 (1.4)	1.5 (1.5)	1.5 (1.5)
Y	3.2 (2.8)	1.5 (1.0)	2.1 (2.0)	2.2 (2.2)	2.2 (2.2)
Z	0.0 (0.0)	1.0 (0.2)	1.1 (0.3)	1.8 (0.5)	1.5 (0.4)
***Trans* Products**	**M1**	0.9 (1.6)	0.2 (0.3)	0.3 (0.4)	0.2 (0.4)
	M2	1.7 (2.3)	0.3 (0.5)	0.4 (0.6)	0.4 (0.6)

(d) + *trans*-**T1** + *trans*-**T2**

Exchange Pool Components (mM)		A	B	C	D
		2.1* (1.7)*	2.7* (2.6)*	1.0* (1.1)*	1.4* (1.4)*
W	1.7 (1.8)	1.4 (1.1)	2.6 (2.7)	3.1 (3.4)	2.9 (3.1)
X	6.1 (6.2)	0.8 (0.6)	1.4 (1.5)	1.5 (1.7)	1.5 (1.6)
Y	3.4 (3.4)	1.3 (1.0)	2.1 (2.2)	2.2 (2.4)	2.2 (2.4)
Z	0.0 (0.0)	0.8 (0.2)	0.8 (0.3)	1.6 (0.5)	1.4 (0.4)
***Trans* Products**	**M1**	1.8 (2.7)	0.2 (0.3)	0.2 (0.4)	0.2 (0.4)
	M2	1.9 (2.8)	0.3 (0.4)	0.4 (0.6)	0.4 (0.6)

A3.3 M1–M3 dynamic covalent libraries

Table 9.8 Raw data describing the distribution of the exchange and product pool components in the **M1**–**M3** experiment after two (blue) and seven (red) days for **(a)** native, **(b)** 10 mol% of *trans*-**T1** **(c)** 10 mol% of *trans*-**T2** and **(d)** 10 mol% of *trans*-**T1** and *trans*-**T2**. Concentrations were quantitatively determined by $^{19}F\{^{1}H\}$ NMR spectroscopy relative to 1-bromo-2-fluoro-4-nitrobenzene. *Trans* product concentrations are corrected for the template added. **Cis* products are not subtracted.

(a) No template

Exchange Pool Components (mM)		A	B	C	D
		2.5* (1.9)*	2.4* (2.8)*	0.8* (1.3)*	1.3* (1.7)*
W	1.5 (1.4)	1.6 (1.2)	2.4 (2.5)	2.9 (3.1)	2.8 (3.0)
X	5.7 (6.0)	0.9 (0.7)	1.3 (1.4)	1.4 (1.6)	1.4 (1.5)
Y	3.1 (3.0)	1.4 (1.0)	2.0 (2.0)	2.1 (2.3)	2.1 (2.2)
Z	0.0 (0.0)	1.2 (0.3)	1.3 (0.5)	2.1 (0.7)	1.8 (0.6)
***Trans* Products**	**M1**	1.9 (4.1)	0.2 (0.4)	0.3 (0.6)	0.3 (0.6)
	M3	0.4 (0.7)	0.3 (0.4)	0.2 (0.4)	0.2 (0.4)

(b) + *trans*-**T1**

Exchange Pool Components (mM)		A	B	C	D
		1.6* (1.2)*	2.5* (2.8)*	0.9* (1.5)*	1.2* (1.7)*
W	1.6 (1.5)	1.5 (1.1)	2.5 (2.5)	3.1 (3.2)	2.9 (3.0)
X	5.9 (5.7)	0.8 (0.6)	1.4 (1.5)	1.5 (1.4)	1.5 (1.6)
Y	3.2 (3.1)	1.3 (0.9)	2.1 (2.1)	2.2 (2.3)	2.2 (2.3)
Z	0.0 (0.0)	0.9 (0.2)	1.2 (0.4)	1.8 (0.6)	1.6 (0.5)
***Trans* Products**	**M1**	3.5 (4.6)	0.2 (0.4)	0.3 (0.6)	0.3 (0.5)
	M3	0.4 (0.6)	0.2 (0.3)	0.2 (0.4)	0.2 (0.3)

(c) + *trans*-**T2**

Exchange Pool Components (mM)		A	B	C	D
		2.2* (2.3)*	2.3* (3.0)*	0.7* (1.4)*	1.2* (1.7)*
W	1.5 (1.4)	1.8 (1.3)	2.5 (2.4)	3.0 (3.1)	2.8 (3.0)
X	5.7 (5.6)	1.0 (0.6)	1.4 (1.4)	1.5 (1.6)	1.5 (1.6)
Y	3.0 (2.9)	1.5 (1.1)	2.0 (2.0)	2.2 (2.2)	2.1 (2.2)
Z	0.0 (0.0)	1.2 (0.3)	1.3 (0.4)	2.1 (0.7)	1.8 (0.6)
***Trans* Products**	**M1**	2.0 (3.8)	0.2 (0.4)	0.3 (0.6)	0.3 (0.6)
	M3	0.3 (0.7)	0.3 (0.3)	0.3 (0.4)	0.3 (0.4)

(d) + *trans*-**T1** + *trans*-**T2**

Exchange Pool Components (mM)		A	B	C	D
		1.9* (2.0)*	2.5* (3.2)*	1.1* (1.7)*	1.4* (2.1)*
W	1.6 (1.5)	1.5 (1.1)	2.6 (2.5)	3.1 (3.1)	2.9 (3.0)
X	6.0 (5.6)	0.8 (0.6)	1.3 (1.3)	1.4 (1.5)	1.4 (1.4)
Y	3.4 (3.1)	1.3 (0.9)	2.0 (2.0)	2.1 (2.2)	2.2 (2.2)
Z	0.0 (0.0)	0.9 (0.2)	1.2 (0.4)	1.8 (0.6)	1.6 (0.5)
***Trans* Products**	**M1**	3.3 (4.7)	0.2 (0.4)	0.3 (0.5)	0.3 (0.5)
	M3	0.4 (0.5)	0.2 (0.3)	0.2 (0.4)	0.2 (0.4)

A3.4 M2–M3 dynamic covalent libraries

Table 9.9 Raw data describing the distribution of the exchange and product pool components in the **M2**–**M3** experiment after two (blue) and seven (red) days for **(a)** native, **(b)** 10 mol% of *trans*-**T1** **(c)** 10 mol% of *trans*-**T2** and **(d)** 10 mol% of *trans*-**T1** and *trans*-**T2**. Concentrations were quantitatively determined by $^{19}F\{^{1}H\}$ NMR spectroscopy relative to 1-bromo-2-fluoro-4-nitrobenzene. *Trans* product concentrations are corrected for the template added. **Cis* products are not subtracted.

(a) No template

Exchange Pool Components (mM)		A	B	C	D
		2.9* (3.4)*	2.4* (2.9)*	0.9* (1.5)*	1.2* (1.7)*
W	1.6 (1.5)	1.8 (1.4)	2.4 (2.5)	2.9 (3.1)	2.8 (2.9)
X	5.8 (5.6)	1.0 (0.8)	1.4 (1.4)	1.4 (1.5)	1.4 (1.5)
Y	3.1 (3.1)	1.5 (1.2)	2.0 (2.0)	2.1 (2.1)	2.1 (2.1)
Z	0.0 (0.0)	1.2 (0.3)	1.3 (0.4)	2.0 (0.6)	1.7 (0.5)
***Trans* Products**	**M2**	1.3 (2.3)	0.3 (0.5)	0.4 (0.7)	0.4 (0.7)
	M3	0.3 (0.6)	0.2 (0.4)	0.3 (0.5)	0.3 (0.5)

(b) + *trans*-**T1**

Exchange Pool Components (mM)		A	B	C	D
		2.7* (3.2)*	2.5* (2.9)*	1.1* (1.5)*	1.3* (1.8)*
W	1.6 (1.5)	1.7 (1.3)	2.5 (2.5)	3.0 (3.2)	2.9 (3.0)
X	5.9 (5.8)	0.9 (0.7)	1.3 (1.3)	1.4 (1.5)	1.4 (1.5)
Y	3.3 (3.1)	1.4 (1.1)	2.0 (2.0)	2.1 (2.2)	2.1 (2.2)
Z	0.0 (0.0)	1.0 (0.2)	1.1 (0.4)	1.8 (0.5)	1.5 (0.5)
***Trans* Products**	**M2**	2.1 (3.1)	0.3 (0.4)	0.4 (0.6)	0.4 (0.6)
	M3	0.3 (0.5)	0.2 (0.4)	0.3 (0.5)	0.3 (0.5)

(c) + *trans*-**T2**

Exchange Pool Components (mM)		A	B	C	D
		2.6* (3.0)*	2.5* (3.0)*	0.9* (1.4)*	1.3* (1.6)*
W	1.6 (1.3)	1.7 (1.3)	2.5 (2.5)	3.0 (3.1)	2.8 (3.1)
X	5.5 (5.3)	0.9 (0.7)	1.4 (1.4)	1.5 (1.5)	1.5 (1.5)
Y	3.0 (3.0)	1.4 (1.1)	2.0 (2.0)	2.2 (2.2)	2.2 (2.2)
Z	0.0 (0.0)	1.0 (0.3)	1.1 (0.4)	1.8 (0.6)	1.5 (0.5)
***Trans* Products**	**M2**	2.2 (3.1)	0.3 (0.5)	0.4 (0.6)	0.4 (0.5)
	M3	0.3 (0.5)	0.2 (0.3)	0.3 (0.5)	0.3 (0.5)

(d) + *trans*-**T1** + *trans*-**T2**

Exchange Pool Components (mM)		A	B	C	D
		2.4* (2.9)*	2.4* (2.9)*	1.0* (1.7)*	1.2* (1.9)*
W	1.7 (1.5)	1.6 (1.2)	2.6 (2.5)	3.1 (3.2)	3.0 (3.0)
X	6.1 (6.2)	0.8 (0.7)	1.3 (1.3)	1.4 (1.5)	1.4 (1.5)
Y	3.4 (3.2)	1.3 (1.0)	2.0 (2.0)	2.2 (2.2)	2.2 (2.1)
Z	0.0 (0.0)	0.9 (0.3)	1.1 (0.5)	1.7 (0.4)	1.5 (0.5)
***Trans* Products**	**M2**	2.7 (3.6)	0.3 (0.4)	0.4 (0.6)	0.4 (0.6)
	M3	0.3 (0.4)	0.2 (0.4)	0.3 (0.5)	0.3 (0.5)

A4. SimFit input file for DCL simulation

```
*================================================================================
* T1T2 simulation, competition of M1 and M2 for AZ
* Condition 2b (EMthermodynamic and EMkinetic = 1)
*================================================================================

MODE (ISOSIM)
* Exchange pool reactions
REACTION ( A + W ==> AW ,  1.00E+4 , 25  )
REACTION ( B + W ==> BW ,  1.00E+4 , 22  )
REACTION ( C + W ==> CW ,  1.00E+4 , 11  )
REACTION ( D + W ==> DW ,  1.00E+4 , 14  )

REACTION ( A + X ==> AX ,  1.00E+4 , 170 )
REACTION ( B + X ==> BX ,  1.00E+4 , 140 )
REACTION ( C + X ==> CX ,  1.00E+4 , 80  )
REACTION ( D + X ==> DX ,  1.00E+4 , 85  )

REACTION ( A + Y ==> AY ,  1.00E+4 , 70  )
REACTION ( B + Y ==> BY ,  1.00E+4 , 60  )
REACTION ( C + Y ==> CY ,  1.00E+4 , 40  )
REACTION ( D + Y ==> DY ,  1.00E+4 , 40  )

REACTION ( A + Z ==> AZ ,  1.00E+4 , 6   )
REACTION ( B + Z ==> BZ ,  1.00E+4 , 7   )
REACTION ( C + Z ==> CZ ,  1.00E+4 , 2.5 )
REACTION ( D + Z ==> DZ ,  1.00E+4 , 3.5 )

* REACTIONS WITH PHENYLACETIC ACID MALEIMIDE M1 and M2

* Bimolecular routes to T1
REACTION (AZ + M1           --> T1          , 1.0000E-4   )
REACTION (AZ + M1           --> cisT1       , 0.3300E-4   )

* Bimolecular routes to T2
REACTION (AZ +  M2          --> T2          , 1.5000E-4   )
REACTION (AZ +  M2          --> cisT2       , 0.7500E-4   )

* Formation of binary complexes including product duplexes
* Only [transT1*transT1] is stable beyond one Pyr*COOH association
REACTION (AZ +  T1           ==> AZT1                    , 1E9,  1.0000E+6   )
REACTION (AZ +  T2           ==> AZT2                    , 1E9,  3.3333E+5   )

REACTION (M1 + T1            ==> M1T1                    , 1E9,  1.0000E+6   )
REACTION (M1 + T2            ==> M1T2                    , 1E9,  1.0000E+6   )
REACTION (M2 + T2            ==> M2T2                    , 1E9,  3.3333E+5   )
REACTION (M2 + T1            ==> M2T1                    , 1E9,  3.3333E+5   )

REACTION (T1 + T1            ==> T1T1                    , 1E9,  33.33333333 )
REACTION (T2 + T2            ==> T2T2                    , 1E9,  3.703703703 )
REACTION (T1 + T2            ==> T1T2                    , 1E9,  11.11111111 )

REACTION (AZ + M1            ==> AZM1                    , 1E9,  1.0000E+6   )
REACTION (AZ + M2            ==> AZM2                    , 1E9,  3.3333E+5   )
REACTION (AZ + T1            ==> AZT1                    , 1E9,  1.0000E+6   )
```

```
REACTION (AZ + T2           ==> AZT2                    , 1E9,  3.3333E+5   )

REACTION (AZ + M1T1         ==> AZM1T1                  , 1E9,  1.0000E+6   )
REACTION (AZ + M1T2         ==> AZM1T2                  , 1E9,  3.3333E+5   )
REACTION (AZ + M2T2         ==> AZM2T2                  , 1E9,  3.3333E+5   )
REACTION (AZ + M2T1         ==> AZM2T1                  , 1E9,  1.0000E+6   )

REACTION (M1 + AZT1         ==> AZM1T1                  , 1E9,  1.0000E+6   )
REACTION (M1 + AZT2         ==> AZM1T2                  , 1E9,  1.0000E+6   )
REACTION (M2 + AZT1         ==> AZM2T1                  , 1E9,  3.3333E+5   )
REACTION (M2 + AZT2         ==> AZM2T2                  , 1E9,  3.3333E+5   )

REACTION (AZ + cisT1        ==> AZcisT1                 , 1E9,  1.0000E+6   )
REACTION (AZ + cisT2        ==> AZcisT2                 , 1E9,  3.3333E+5   )
REACTION (M1 + cisT1        ==> M1cisT1                 , 1E9,  1.0000E+6   )
REACTION (M1 + AZcisT1      ==> AZM1cisT1               , 1E9,  1.0000E+6   )
REACTION (M2 + cisT2        ==> M2cisT2                 , 1E9,  3.3333E+5   )
REACTION (M2 + AZcisT2      ==> AZM2cisT2               , 1E9,  3.3333E+5   )

REACTION (M2 + cisT1        ==> M2cisT1                 , 1E9,  3.3333E+5   )
REACTION (M2 + AZcisT1      ==> AZM2cisT1               , 1E9,  3.3333E+5   )
REACTION (M1 + cisT2        ==> M1cisT2                 , 1E9,  1.0000E+6   )
REACTION (M1 + AZcisT2      ==> AZM1cisT2               , 1E9,  1.0000E+6   )

REACTION (AZ + M1cisT1      ==> AZM1cisT1               , 1E9,  1.0000E+6   )
REACTION (AZ + M2cisT2      ==> AZM2cisT2               , 1E9,  3.3333E+5   )
REACTION (AZ + M2cisT1      ==> AZM2cisT1               , 1E9,  1.0000E+6   )
REACTION (AZ + M1cisT2      ==> AZM1cisT2               , 1E9,  3.3333E+5   )

REACTION (cisT1  + cisT1    ==> cisT1cisT1              , 1E9,  1.0000E+6   )
REACTION (cisT2  + cisT2    ==> cisT2cisT2              , 1E9,  3.3333E+5   )
REACTION (cisT2  + cisT1    ==> cisT2cisT2              , 1E9,  577350.2691 )

REACTION (T1 + cisT1        ==> T1cisT1                 , 1E9,  1.0000E+6   )
REACTION (T2 + cisT2        ==> T2cisT2                 , 1E9,  3.3333E+5   )
REACTION (T2 + cisT1        ==> T2cisT1                 , 1E9,  577350.2691 )
REACTION (T1 + cisT2        ==> T1cisT2                 , 1E9,  577350.2691 )

* Ternary complex reaction
REACTION (AZM1T1      --> T1T1          , 100.00E-4       )
REACTION (AZM1T2      --> T1T2          , 5.0000E-4       )
REACTION (AZM2T2      --> T2T2          , 150.000E-4      )
REACTION (AZM2T1      --> T1T2          , 15.000E-4       )

*Bimolecular Reactions of Complexes
REACTION (AZM1 + AZ         --> T1 + AZ                       , 1.0000E-4   )
REACTION (AZM1 + M1         --> T1 + M1                       , 1.0000E-4   )
REACTION (AZM1 + AZ         --> cisT1 + AZ                    , 0.3300E-4   )
REACTION (AZM1 + M1         --> cisT1 + M1                    , 0.3300E-4   )

REACTION (AZM2 +  AZ        --> T2 +  AZ                      , 1.5000E-4   )
REACTION (AZM2 +  M2        --> T2 +  M2                      , 1.5000E-4   )
REACTION (AZM2 +  AZ        --> cisT2 +  AZ                   , 0.7500E-4   )
REACTION (AZM2 +  M2        --> cisT2 +  M2                   , 0.7500E-4   )

REACTION (AZM1 +  M2        --> T1 + M2                       , 1.0000E-4   )
REACTION (AZM1 +  M2        --> cisT1 + M2                    , 0.3300E-4   )
```

```
REACTION (AZM2 +  M1         --> T2 +  M1                      , 1.5000E-4   )
REACTION (AZM2 +  M1         --> cisT2 +  M1                   , 0.7500E-4   )

REACTION (AZM2 +  M1         --> T2 + M1                       , 1.5000E-4   )
REACTION (AZM2 +  M1         --> cisT2 + M1                    , 0.7500E-4   )
REACTION (AZM1 +  M2         --> T1 +  M2                      , 1.0000E-4   )
REACTION (AZM1 +  M2         --> cisT1 +  M2                   , 0.3300E-4   )

REACTION (M1T1 + AZ          --> T1 + T1                       , 1.0000E-4   )
REACTION (M1T1 + AZ          --> cisT1 + T1                    , 0.3300E-4   )
REACTION (M2T1 + AZ          --> T2 + T1                       , 1.5000E-4   )
REACTION (M2T1 + AZ          --> cisT2 + T1                    , 0.7500E-4   )

REACTION (M1T2 + AZ          --> T1 + T2                       , 1.0000E-4   )
REACTION (M1T2 + AZ          --> cisT1 + T2                    , 0.3300E-4   )
REACTION (M2T2 + AZ          --> T2 + T2                       , 1.5000E-4   )
REACTION (M2T2 + AZ          --> cisT2 + T2                    , 0.7500E-4   )

REACTION (AZM1 + AZcisT1     --> T1 + AZ + cisT1               , 1.0000E-4   )
REACTION (AZM1 + M1cisT1     --> T1 + M1 + cisT1               , 1.0000E-4   )
REACTION (AZM1 + AZT1        --> T1 + AZ + T1                  , 1.0000E-4   )
REACTION (AZM1 + M1T1        --> T1 + M1 + T1                  , 1.0000E-4   )
REACTION (AZM1 + AZT2        --> T1 + AZ + T2                  , 1.0000E-4   )
REACTION (AZM1 + M1T2        --> T1 + M1 + T2                  , 1.0000E-4   )

REACTION (AZM2 +  AZcisT2    --> T2 +  AZ + cisT2              , 1.5000E-4   )
REACTION (AZM2 +  M2cisT2    --> T2 +  M2 + cisT2              , 1.5000E-4   )
REACTION (AZM2 +  AZT2       --> T2 +  AZ + T2                 , 1.5000E-4   )
REACTION (AZM2 +  M2T2       --> T2 +  M2 + T2                 , 1.5000E-4   )
REACTION (AZM2 +  AZT1       --> T2 +  AZ + T1                 , 1.5000E-4   )
REACTION (AZM2 +  M2T1       --> T2 +  M2 + T1                 , 1.5000E-4   )

REACTION (AZM1 + AZcisT1     --> cisT1  + AZ + cisT1           , 0.3300E-4   )
REACTION (AZM1 + M1cisT1     --> cisT1  + M1 + cisT1           , 0.3300E-4   )
REACTION (AZM1 + AZT1        --> cisT1  + AZ + T1              , 0.3300E-4   )
REACTION (AZM1 + M1T1        --> cisT1  + M1 + T1              , 0.3300E-4   )
REACTION (AZM1 + AZT2        --> cisT1  + AZ + T2              , 0.3300E-4   )
REACTION (AZM1 + M1T2        --> cisT1  + M1 + T2              , 0.3300E-4   )

REACTION (AZM2 +  AZcisT2    --> cisT2  +  AZ + cisT2          , 0.7500E-4   )
REACTION (AZM2 +  M2cisT2    --> cisT2  +  M2 + cisT2          , 0.7500E-4   )
REACTION (AZM2 +  AZT2       --> cisT2  +  AZ + T2             , 0.7500E-4   )
REACTION (AZM2 +  M2T2       --> cisT2  +  M2 + T2             , 0.7500E-4   )
REACTION (AZM2 +  AZT1       --> cisT2  +  AZ + T1             , 0.7500E-4   )
REACTION (AZM2 +  M2T1       --> cisT2  +  M2 + T1             , 0.7500E-4   )

REACTION (AZM1 + AZM1cisT1   --> cisT1  + AZ + M1 + cisT1      , 0.3300E-4   )
REACTION (AZM1 + AZM1cisT1   --> T1 + AZ + M1 + cisT1          , 1.0000E-4   )
REACTION (AZM1 + AZM1T1      --> cisT1  + AZ + M1 + T1         , 0.3300E-4   )
REACTION (AZM1 + AZM1T1      --> T1 + AZ + M1 + T1             , 1.0000E-4   )
REACTION (AZM1 + AZM1T2      --> cisT1  + AZ + M1 + T2         , 0.3300E-4   )
REACTION (AZM1 + AZM1T2      --> T1 + AZ + M1 + T2             , 1.0000E-4   )

REACTION (AZM2 +  AZM2cisT2  --> cisT2   +  AZ +  M2 + cisT2   , 0.7500E-4   )
REACTION (AZM2 +  AZM2cisT2  --> T2 +  AZ +  M2 + cisT2        , 1.5000E-4   )
REACTION (AZM2 +  AZM2T2     --> cisT2   +  AZ +  M2 + T2      , 0.7500E-4   )
REACTION (AZM2 +  AZM2T2     --> T2 +  AZ +  M2 + T2           , 1.5000E-4   )
```

```
REACTION (AZM2 +  AZM2T1     --> cisT2   +  AZ +  M2 + T1        , 0.7500E-4   )
REACTION (AZM2 +  AZM2T1     --> T2 +  AZ +  M2 + T1             , 1.5000E-4   )

REACTION (       2 AZM1      --> T1 + AZ + M1                    , 1.0000E-4   )
REACTION (       2 AZM1      --> cisT1  + AZ + M1                , 0.3300E-4   )
REACTION (       2 AZM2      --> T2 +  AZ +  M2                  , 1.5000E-4   )
REACTION (       2 AZM2      --> cisT2   +  AZ +  M2             , 0.7500E-4   )

* Reactions with other, recognition-disabled nitrones BZ, CZ and DZ
REACTION (BZ + M1           --> transT1BZ    , 1.0000E-4 )
REACTION (BZ + M1           --> cisT1BZ      , 0.3300E-4 )
REACTION (BZ + M2           --> transT2BZ    , 1.5000E-4 )
REACTION (BZ + M2           --> cisT2BZ      , 0.7500E-4 )

REACTION (CZ + M1           --> transT1CZ    , 1.0000E-4 )
REACTION (CZ + M1           --> cisT1CZ      , 0.3300E-4 )
REACTION (CZ + M2           --> transT2CZ    , 1.5000E-4 )
REACTION (CZ + M2           --> cisT2CZ      , 0.7500E-4 )

REACTION (DZ + M1           --> transT1DZ    , 1.0000E-4          )
REACTION (DZ + M1           --> cisT1DZ      , 0.3300E-4          )
REACTION (DZ + M2           --> transT2DZ    , 1.5000E-4          )
REACTION (DZ + M2           --> cisT2DZ      , 0.7500E-4          )

* Association of A  with M1 and M2, and T1 and T2 and their bimolecular reactions
REACTION (AW + M1            ==> AWM1         , 1E9,  1.0000E+6   )
REACTION (AW + M2            ==> AWM2         , 1E9,  3.3333E+5   )
REACTION (AW + T1            ==> AWT1         , 1E9,  1.0000E+6   )
REACTION (AW + T2            ==> AWT2         , 1E9,  3.3333E+5   )

REACTION (AZ + AWM1          --> T1    + AW   , 1.0000E-4         )
REACTION (AZ + AWM1          --> cisT1 + AW   , 0.3300E-4         )
REACTION (AZ + AWM2          --> T2    + AW   , 1.5000E-4         )
REACTION (AZ + AWM2          --> cisT2 + AW   , 0.7500E-4         )

REACTION (AX + M1            ==> AXM1         , 1E9,  1.0000E+6   )
REACTION (AX + M2            ==> AXM2         , 1E9,  3.3333E+5   )
REACTION (AX + T1            ==> AXT1         , 1E9,  1.0000E+6   )
REACTION (AX + T2            ==> AXT2         , 1E9,  3.3333E+5   )

REACTION (AZ + AXM1          --> T1    + AX   , 1.0000E-4         )
REACTION (AZ + AXM1          --> cisT1 + AX   , 0.3300E-4         )
REACTION (AZ + AXM2          --> T2    + AX   , 1.5000E-4         )
REACTION (AZ + AXM2          --> cisT2 + AX   , 0.7500E-4         )

REACTION (AY + M1            ==> AYM1         , 1E9,  1.0000E+6   )
REACTION (AY + M2            ==> AYM2         , 1E9,  3.3333E+5   )
REACTION (AY + T1            ==> AYT1         , 1E9,  1.0000E+6   )
REACTION (AY + T2            ==> AYT2         , 1E9,  3.3333E+5   )

REACTION (AZ + AYM1          --> T1    + AY   , 1.0000E-4         )
REACTION (AZ + AYM1          --> cisT1 + AY   , 0.3300E-4         )
REACTION (AZ + AYM2          --> T2    + AY   , 1.5000E-4         )
REACTION (AZ + AYM2          --> cisT2 + AY   , 0.7500E-4         )

REACTION ( COMPILE )
REACTION ( SHOW    )
```

```
CONSTANT ( SHOW    )

*INITIALISE COCNENTRATIONS
   INIT     (  A  , 0.05 , 1 )
   INIT     (  B  , 0.05 , 2 )
   INIT     (  C  , 0.05 , 3 )
   INIT     (  D  , 0.05 , 4 )
   INIT     (  W  , 0.05 , 5 )
   INIT     (  X  , 0.05 , 6 )
   INIT     (  Y  , 0.05 , 7 )
   INIT     (  Z  , 0.05 , 8 )
   INIT     ( M1  , 0.05 , 9 )
   INIT     ( M2  , 0.05 , 10)

INTEG    ( STIFF, 1E-10, 50, 0.025, 100, 50 )
NUMPLOT  ( 100  )
TIME     ( SEC  )
WIN      ( 0, 172800, 250, 0.1, 0, 10E-3, 5E-3, 3E-4 )
PLOT     ( FILE )
PLOT
```

A5. Example SimFit files employed in rotaxane simulations

A5.1 SimFit input file for PM-R simulations

```
*================================================================================
* Simulation for formation of PMR at different Ka [LM]
* Ka for individual binding event is 910 M-1, hence K off 1.1 x 10^6
* L = Nitrone and S = Maleimide, M = macrocycle
*================================================================================

   MODE     ( ISOSIM )
*Formation of [L-M] complex
REACTION (L + M           ==> LM          ,1.00E+9  ,1E+4 )

* Bimolecular routes to THREAD
REACTION (L + S           --> THREAD      ,1.0327E-4      )
REACTION (LM + S          --> PMR         ,1.0327E-4      )

* Bimolecular routes to Rotaxane
REACTION (L + S           --> CIS         ,0.3440E-4      )
REACTION (LM + S          --> CISR        ,0.3440E-4      )

* Formation of binary complexes including product duplexes
* Only [TEMPLATE*TEMPLATE] is stable beyond one Pyr*COOH association
REACTION (L + THREAD       ==> LTHREAD       ,1.00E+9  ,1.1E+6    )
REACTION (LM + THREAD      ==> LMTHREAD      ,1.00E+9  ,1.1E+6    )
REACTION (L + PMR          ==> LPMR          ,1.00E+9  ,1.1E+6    )
REACTION (LM + PMR         ==> LMPMR         ,1.00E+9  ,1.1E+6    )
REACTION (S + THREAD       ==> STHREAD       ,1.00E+9  ,1.1E+6    )
REACTION (S + PMR          ==> SPMR          ,1.00E+9  ,1.1E+6    )
REACTION (L + STHREAD      ==> LSTHREAD      ,1.00E+9  ,1.1E+6    )
REACTION (LM + STHREAD     ==> LMSTHREAD     ,1.00E+9  ,1.1E+6    )
REACTION (L + SPMR         ==> LSPMR         ,1.00E+9  ,1.1E+6    )
REACTION (LM + SPMR        ==> LMSPMR        ,1.00E+9  ,1.1E+6    )
REACTION (S + LTHREAD      ==> LSTHREAD      ,1.00E+9  ,1.1E+6    )
REACTION (S + LMTHREAD     ==> LMSTHREAD     ,1.00E+9  ,1.1E+6    )
REACTION (S + LPMR         ==> LSPMR         ,1.00E+9  ,1.1E+6    )
REACTION (S + LMPMR        ==> LMSPMR        ,1.00E+9  ,1.1E+6    )
REACTION (THREAD + THREAD  ==> THREADTHREAD  ,1.00E+9  ,9.8291E+1 )
REACTION (THREAD + PMR     ==> THREADPMR     ,1.00E+9  ,9.8291E+1 )
REACTION (PMR + PMR        ==> PMRPMR        ,1.00E+9  ,9.8291E+1 )
REACTION (L + S            ==> LS            ,1.00E+9  ,1.1E+6    )
REACTION (LM + S           ==> LMS           ,1.00E+9  ,1.1E+6    )

* Ternary complex reaction
REACTION (LSTHREAD      --> THREADTHREAD           ,2.2587E-2 )
REACTION (LMSTHREAD     --> THREADPMR              ,2.2587E-2 )
REACTION (LSPMR         --> THREADPMR              ,2.2587E-2 )
REACTION (LMSPMR        --> PMRPMR                 ,2.2587E-2 )

*Bimolecular Reactions of Complexes
REACTION (LS + L       --> THREAD + L              ,1.0327E-4 )
REACTION (LMS + L      --> PMR    + L              ,1.0327E-4 )
```

```
REACTION (LS + S       --> THREAD + S               ,1.0327E-4 )
REACTION (LMS + S      --> PMR    + S               ,1.0327E-4 )

REACTION (LS + LTHREAD  --> THREAD + L + THREAD     ,1.0327E-4 )
REACTION (LMS + LTHREAD --> PMR + L + THREAD        ,1.0327E-4 )
REACTION (LS + STHREAD  --> THREAD + S + THREAD     ,1.0327E-4 )
REACTION (LMS + STHREAD --> PMR + S + THREAD        ,1.0327E-4 )
REACTION (LS + LPMR     --> THREAD + L + PMR        ,1.0327E-4 )
REACTION (LMS + LPMR    --> PMR + L + PMR           ,1.0327E-4 )
REACTION (LS + SPMR     --> THREAD + S + PMR        ,1.0327E-4 )
REACTION (LMS + SPMR    --> PMR + S + PMR           ,1.0327E-4 )

REACTION (LS + LSTHREAD  --> THREAD + L + S + THREAD,1.0327E-4 )
REACTION (LMS + LSTHREAD --> PMR + L + S + THREAD   ,1.0327E-4 )
REACTION (LS + LSPMR     --> THREAD + L + S + PMR   ,1.0327E-4 )
REACTION (LMS + LSPMR    --> PMR + L + S + PMR      ,1.0327E-4 )

REACTION (       2 LS  --> THREAD + L + S           ,1.0327E-4 )
REACTION (       2 LMS --> PMR + LM + S             ,1.0327E-4 )

*REACTION (L + M                 ==> NitroneM          ,1.00E+9  ,1E+6 )
*REACTION (L + M                 ==> AmidopyridineM    ,1.00E+9  ,1E+6 )
*REACTION (AmidopyridineM + S    --> THREAD + M        ,1.0327E-4      )
*REACTION (THREAD + M            ==> THREADM           ,1.00E+9  ,1E+6 )
*REACTION (PMR + M               ==> PMRM              ,1.00E+9  ,1E+6 )

REACTION ( COMPILE )
REACTION ( SHOW    )

INIT   (L, 0.01, 1) SCALE (3,1)
INIT   (M, 0.01, 2) SCALE (3,1)
INIT   (S, 0.01, 3) SCALE (3,1)

SINTEG    (STIFF, 1E-9, 8, 0.05, 200, 100)

NUMPLOT  ( 10   )
TIME     ( SEC  )
WIN      ( 0, 28800, 25000, 0.1, 0, 20E-3, 5E-3, 3E-4 )
PLOT     ( FILE )
PLOT
```

A5.2 SimFit input file for Model 3 simulations

```
*================================================================================
* Model 3 Rotaxane simulation (Condition Ia)
* Ka for individual binding event is 910 M-1, hence K off 1.1 x 10^6
* L = Nitrone and S = Maleimide, M = macrocycle, PMR = rotaxane
* EMkinetic = 10 and EMthermo = 0.1
*================================================================================

   MODE     ( ISOSIM )
*Formation of [L-M] complex
REACTION (L + M            ==> LM          ,1.00E+9  ,5E+6 )
REACTION (L + MSPMR        ==> LMSPMR      ,1.00E+9  ,5E+6 )
REACTION (L + MS           ==> LMS         ,1.00E+9  ,5E+6 )
REACTION (L + MPMR         ==> LMPMR       ,1.00E+9  ,5E+6 )
REACTION (L + MTHREAD      ==> LMTHREAD    ,1.00E+9  ,5E+6 )

* Bimolecular routes to THREAD and ROTAXANE
REACTION (L  + S           --> THREAD      ,1.0327E-4      )
REACTION (LM + S           --> PMR         ,1.0327E-4      )
REACTION (MS + L           --> THREAD + M  ,1.0327E-4      )
REACTION (L  + S           --> CIS         ,0.3440E-4      )
REACTION (LM + S           --> CISR        ,0.3440E-4      )

* Formation of binary complexes including product duplexes
* Only [TEMPLATE*TEMPLATE] is stable beyond one Pyr*COOH association
REACTION (LM  + S            ==> LMS           ,1.00E+9  ,1.1E+6    )
REACTION (M   + S            ==> MS            ,1.00E+9  ,1.1E+6    )
REACTION (M   + PMR          ==> MPMR          ,1.00E+9  ,1.1E+6    )
REACTION (M   + THREAD       ==> MTHREAD       ,1.00E+9  ,1.1E+6    )
REACTION (LM  + THREAD       ==> LMTHREAD      ,1.00E+9  ,1.1E+6    )
REACTION (LM  + PMR          ==> LMPMR         ,1.00E+9  ,1.1E+6    )
REACTION (S   + PMR          ==> SPMR          ,1.00E+9  ,1.1E+6    )
REACTION (LM  + SPMR         ==> LMSPMR        ,1.00E+9  ,1.1E+6    )
REACTION (S   + LMPMR        ==> LMSPMR        ,1.00E+9  ,1.1E+6    )

REACTION (THREAD + PMR       ==> THREADPMR     ,1.00E+9  ,1.1E+6    )
REACTION (PMR    + PMR       ==> PMRPMR        ,1.00E+9  ,1.2076E+4 )

* Ternary complex reaction

REACTION (LMSPMR         --> PMRPMR                  ,10.3274E-4   )

*Bimolecular Reactions of Complexes

REACTION (LMS + L        --> PMR    + L              ,1.0327E-4    )
REACTION (LMS + S        --> PMR    + S              ,1.0327E-4    )
REACTION (LMS + LM       --> PMR    + LM             ,1.0327E-4    )

REACTION (LMS + SPMR     --> PMR + S + PMR           ,1.0327E-4    )
REACTION (LMS + MSPMR    --> PMR + MS + PMR          ,1.0327E-4    )

REACTION (LMS + LMS      --> PMR + LM + S            ,1.0327E-4    )
REACTION (LM  + SPMR     --> PMR + PMR               ,1.0327E-4    )
REACTION (S   + LMPMR    --> PMR + PMR               ,1.0327E-4    )
REACTION (S   + LMTHREAD --> PMR + THREAD            ,1.0327E-4    )
```

```
REACTION ( COMPILE )
REACTION ( SHOW    )

INIT   (L, 0.05, 1) SCALE (3,1)
INIT   (M, 0.05, 2) SCALE (3,1)
INIT   (S, 0.05, 3) SCALE (3,1)

SINTEG   ( STIFF, 1E-9, 8, 0.05, 200, 100          )
NUMPLOT  ( 100  )
TIME     ( SEC  )
WIN      (0, 28800, 25000, 0.1, 0, 20E-3, 5E-3, 3E-4)
PLOT     ( FILE )
PLOT
```

A6. Estimation of errors in the calculation of % enhancements

Standard rules for calculation of error propagation[a] were employed in the determination of the error values for the % enhancements (**Equation 9.1**) in **Chapter 3**. These rules allow us to take certain quantities (in this case product concentration) associated with a particular uncertainty error ($\delta = \pm\ 0.02$ mM, see **Chapter 9** for details of how these were determined) and use these to calculate another quantity, *i.e*. the % enhancements in this case.

$$\%Enhancement = 100 \cdot \frac{[\mathrm{T}]_{instructed} - [\mathrm{T}]_{uninstructed}}{[\mathrm{T}]_{uninstructed}} \qquad \textbf{(9.1)}$$

In order to determine the uncertainty in the calculated values of these % enhancements in **Chapter 3**, the following rules were employed. The uncertainty associated with the numerator in **Equation 9.1** can be expressed in the following form (**Equation 9.2**):

$$\delta([\mathrm{T}]_{instructed} - [\mathrm{T}]_{uninstructed}) = \sqrt{(\delta[\mathrm{T}]_{instructed})^2 + (\delta[\mathrm{T}]_{uninstructed})^2} \qquad \textbf{(9.2)}$$

where the δ terms for both of these concentrations ($[\mathrm{T}]_{instructed}$ and $[\mathrm{T}]_{uninstructed}$, right-hand side of **Equation 9.2**) are estimated at $\pm\ 0.02$ mM, giving an overall value for the error of the numerator in each case of $\pm\ 0.028$ mM. This value can be used to determine the uncertainty corresponding to the overall % enhancement for each template-instructed catalytic pathways using **Equation 9.3**:

$$\frac{\delta\%Enhancement}{\%Enhancement} = \sqrt{\left(\frac{\delta([\mathrm{T}]_{instructed} - [\mathrm{T}]_{uninstructed})}{[\mathrm{T}]_{instructed} - [\mathrm{T}]_{uninstructed}}\right)^2 + \left(\frac{\delta[\mathrm{T}]_{uninstructed}}{[\mathrm{T}]_{uninstructed}}\right)^2} \qquad \textbf{(9.3)}$$

Here, the terms $\delta([\mathrm{T}]_{instrcuted} - [\mathrm{T}]_{uninstructed})$ ($\pm\ 0.028$ mM) and $\delta[\mathrm{T}]_{uninstrcuted}$ ($\pm\ 0.02$ mM) remain the same in each case, and only the values of concentration vary according to the experiment. The value of each % enhancement and the error (δ) in % enhancement associated with it were rounded up to the nearest whole numbers (see **Chapter 3** for detailed discussion of these values).

[a]Employing the assumption that these errors are normally distributed.